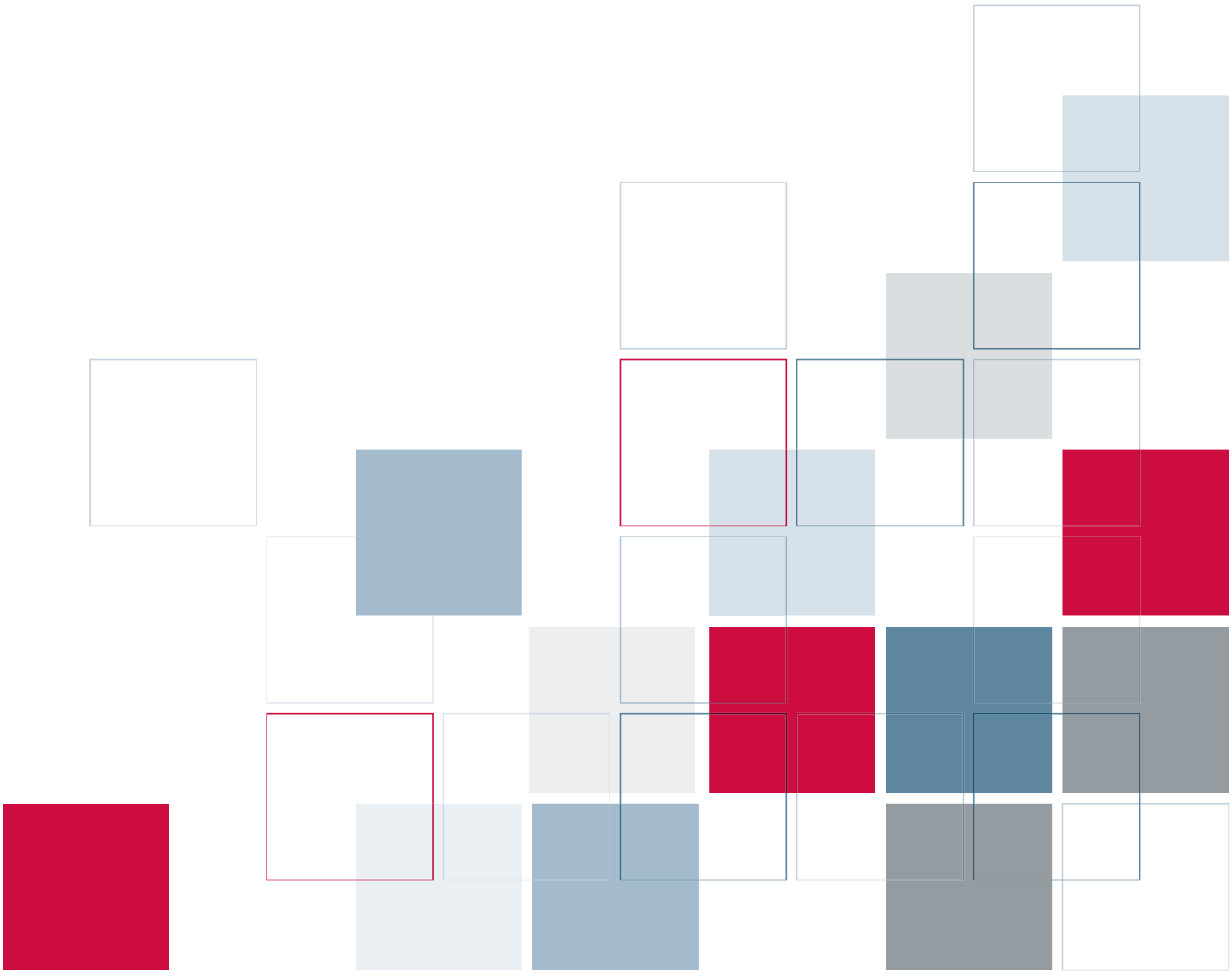
SPSS Trends™ 16.0



Weitere Informationen zu SPSS®-Software-Produkten finden Sie auf unserer Website unter der Adresse *http://www.spss.com* oder wenden Sie sich an

SPSS Inc.
233 South Wacker Drive, 11th Floor
Chicago, IL 60606-6412, USA
Tel.: (312) 651-3000
Fax: (312) 651-3668

SPSS ist eine eingetragene Marke, und weitere Produktnamen sind Marken der SPSS Inc. für Computerprogramme von SPSS Inc. Die Herstellung oder Verbreitung von Materialien, die diese Programme beschreiben, ist ohne die schriftliche Erlaubnis des Eigentümers der Marke und der Lizenzrechte der Software und der Copyrights der veröffentlichten Materialien verboten.

Die SOFTWARE und die Dokumentation werden mit BESCHRÄNKTEN RECHTEN zur Verfügung gestellt. Verwendung, Vervielfältigung und Veröffentlichung durch die Regierung unterliegen den Beschränkungen in Unterabschnitt (c)(1)(ii) von The Rights in Technical Data and Computer Software unter 52.227-7013. Vertragspartner/Hersteller ist SPSS Inc., 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.
Patentnr. 7.023.453

Allgemeiner Hinweis: Andere in diesem Dokument verwendete Produktnamen werden nur zu Identifikationszwecken genannt und können Marken der entsprechenden Unternehmen sein.

Windows ist eine eingetragene Marke der Microsoft Corporation.

Apple, Mac und das Mac-Logo sind Marken von Apple Computer, Inc., die in den USA und in anderen Ländern eingetragen sind.

Dieses Produkt verwendet WinWrap Basic, Copyright 1993–2007, Polar Engineering and Consulting, *http://www.winwrap.com*.

SPSS Trends™ 16.0
Copyright © 2007 SPSS Inc.
Alle Rechte vorbehalten.

Ohne schriftliche Erlaubnis der SPSS GmbH Software darf kein Teil dieses Handbuchs für irgendwelche Zwecke oder in irgendeiner Form mit irgendwelchen Mitteln, elektronisch oder mechanisch, mittels Fotokopie, durch Aufzeichnung oder durch andere Informationsspeicherungssysteme reproduziert werden.

1 2 3 4 5 6 7 8 9 0 10 09 08 07

Vorwort

SPSS 16.0 ist ein umfassendes System zum Analysieren von Daten. Das optionale Erweiterungsmodul SPSS Trends bietet die zusätzlichen Analyseverfahren, die in diesem Handbuch beschrieben sind. Die Prozeduren im Erweiterungsmodul Trends müssen zusammen mit SPSS 16.0 Base verwendet werden. Sie sind vollständig in dieses System integriert.

Installation

Zur Installation von SPSS Trends Erweiterungsmodul führen Sie den Lizenzautorisierungsassistenten mit dem Autorisierungscode aus, den Sie von SPSS erhalten haben. Weitere Informationen finden Sie in den Installationsanweisungen im Lieferumfang von SPSS Trends Erweiterungsmodul.

Kompatibilität

SPSS kann auf vielen Computersystemen ausgeführt werden. Mindestanforderungen an das System und Empfehlungen finden Sie in den Unterlagen, die mit Ihrem System geliefert werden.

Seriennummern

Die Seriennummer des Programms dient gleichzeitig als Identifikationsnummer bei SPSS. Sie benötigen diese Seriennummer, wenn Sie sich an SPSS wenden, um Informationen über Kundendienst, zu Zahlungen oder Aktualisierungen des Systems zu erhalten. Die Seriennummer wird mit dem Base-System ausgeliefert.

Kundendienst

Wenden Sie sich mit Fragen bezüglich der Lieferung oder Ihres Kundenkontos an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Halten Sie bitte stets Ihre Seriennummer bereit.

Ausbildungsseminare

SPSS bietet öffentliche und unternehmensinterne Seminare an. Alle Seminare beinhalten auch praktische Übungen. Seminare finden in größeren Städten regelmäßig statt. Wenn Sie weitere Informationen zu diesen Schulungen wünschen, wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden.

Technischer Support

Kunden von SPSS mit Wartungsvertrag können den Technischen Support in Anspruch nehmen. Kunden können sich an den Technischen Support wenden, wenn sie Hilfe bei der Arbeit mit SPSS oder bei der Installation in einer der unterstützten Hardware-Umgebungen benötigen. Informationen über den Technischen Support finden Sie auf der Website von SPSS unter *http://www.spss.com* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Bei einem Anruf werden Sie nach Ihrem Namen, dem Namen Ihrer Organisation und Ihrer Seriennummer gefragt.

Weitere Veröffentlichungen

Weitere Exemplare von Produkthandbüchern können direkt bei SPSS Inc. bestellt werden. Besuchen Sie den SPSS Web Store unter *http://www.spss.com/estore* oder wenden Sie sich an Ihr regionales SPSS-Büro, das Sie auf der SPSS-Website unter *http://www.spss.com/worldwide* finden. Wenden Sie sich bei telefonischen Bestellungen in den USA und Kanada unter 800-543-2185 direkt an SPSS Inc. Wenden Sie sich bei telefonischen Bestellungen außerhalb von Nordamerika an Ihr regionales SPSS-Büro, das Sie auf der SPPS-Website finden.

Das Handbuch *SPSS Statistical Procedures Companion* von Marija Norušis wurde von Prentice Hall veröffentlicht. Eine neue Fassung dieses Buchs mit Aktualisierungen für SPSS 16.0 ist geplant. Das Handbuch *SPSS Advanced Statistical Procedures Companion*, bei dem auch SPSS 16.0 berücksichtigt wird, erscheint demnächst. Das Handbuch *SPSS Guide to Data Analysis* für SPSS 16.0 wird ebenfalls derzeit erstellt. Ankündigungen für Veröffentlichungen, die ausschließlich über Prentice Hall verfügbar sind, finden Sie auf der SPSS-Website unter *http://www.spss.com/estore* (wählen Sie Ihr Land aus und klicken Sie auf Books).

Kundenmeinungen

Ihre Meinung ist uns wichtig. Teilen Sie uns bitte Ihre Erfahrungen mit SPSS-Produkten mit. Insbesondere haben wir Interesse an neuen, interessanten Anwendungsgebieten von SPSS Trends Erweiterungsmodul. Senden Sie uns eine E-Mail an *suggest@spss.com* oder schreiben Sie an: SPSS Inc., Attn: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Über dieses Handbuch

In diesem Handbuch wird die grafische Benutzeroberfläche für die in SPSS Trends Erweiterungsmodul enthaltenen Prozeduren erläutert. Die Abbildungen der Dialogfelder stammen aus SPSS. Detaillierte Informationen zur Befehlssyntax für die Funktionen in SPSS Trends Erweiterungsmodul sind auf zwei Arten verfügbar: als Bestandteil des umfassenden Hilfesystems und als separates Dokument im PDF-Format im Handbuch *SPSS 16.0 Command Syntax Reference*, das auch über das Menü “Hilfe” verfügbar ist.

Kontakt zu SPSS

Wenn Sie in unseren Verteiler aufgenommen werden möchten, wenden Sie sich an eines unserer Büros, die Sie auf unserer Website unter *http://www.spss.com/worldwide* finden.

Inhalt

Teil I: Benutzerhandbuch

1 Einführung in Zeitreihen 1

Zeitreihendaten 1
Transformieren von Daten. 2
Schätz- und Validierungsperioden 2
Erstellen von Modellen und Vorhersagen 2

2 Zeitreihenmodellierung 4

Angeben von Optionen für den Expert Modeler. 8
Modellauswahl und Ereignisspezifikation. 8
Behandeln von Ausreißern mit dem Expert Modeler. 9
Exponentielles Glätten: Benutzerdefinierte Modelle 10
Benutzerdefinierte ARIMA-Modelle 12
Modellspezifikation für benutzerdefinierte ARIMA-Modelle 12
Übertragungsfunktionen in benutzerdefinierten ARIMA-Modellen 14
Ausreißer in benutzerdefinierten ARIMA-Modellen 16
Ausgabe. 17
Statistiken und Vorhersagetabellen 17
Diagramme 19
Einschränken der Ausgabe auf die am besten/schlechtesten angepassten Modelle 21
Speichern von Modellvorhersagen und Modellspezifikationen 22
Optionen. 24
Zusätzliche Funktionen beim Befehl TSMODEL 26

3 Zeitreihenmodell zuweisen 27

Ausgabe. 30
Statistiken und Vorhersagetabellen 31
Diagramme 33
Einschränken der Ausgabe auf die am besten/schlechtesten angepassten Modelle 35
Speichern von Modellvorhersagen und Modellspezifikationen 37

Optionen. 38
Zusätzliche Funktionen beim Befehl TSAPPLY. 39

4 Saisonale Zerlegung 40

Saisonale Zerlegung: Speichern 42
Zusätzliche Funktionen beim Befehl SEASON 43

5 Spektraldiagramme 44

Zusätzliche Funktionen beim Befehl SPECTRA 47

Teil II: Beispiele

6 Prognosen für große Datenmengen mit dem Expert Modeler 49

Untersuchen der Daten. 49
Durchführung der Analyse 51
Diagramme für die Modellzusammenfassung 57
Modellvorhersagen 58
Zusammenfassung 59

7 Neuprognosen für große Datenmengen durch das Zuweisen gespeicherter Modelle 60

Durchführung der Analyse 60
Statistiken zur Anpassungsgüte des Modells 63
Modellvorhersagen 64
Zusammenfassung 64

8 Ermitteln von signifikanten Einflussvariablen mit dem Expert Modeler 66

Erstellen von Diagrammen aus den Daten . . . 66
Durchführung der Analyse . . . 68
Zeitreihendiagramm . . . 74
Tabelle der Modellbeschreibung . . . 74
Tabelle der Modellstatistik . . . 75
Tabelle der ARIMA-Modellparameter . . . 76
Zusammenfassung . . . 76

9 Untersuchen von Einflussvariablen durch das Zuweisen gespeicherter Modelle 77

Erweitern der Prädiktorzeitreihe . . . 77
Ändern der Werte von Einflussvariablen in der Vorhersageperiode . . . 81
Durchführung der Analyse . . . 83

10 Saisonale Zerlegung 88

Entfernen der Saisonalität aus Verkaufsdaten . . . 88
Bestimmen und Festlegen der Periodizität . . . 88
Durchführung der Analyse . . . 92
Interpretation der Ausgabe . . . 93
Zusammenfassung . . . 95
Verwandte Prozeduren . . . 95

11 Spektraldiagramme 96

Verwendung von Spektraldiagrammen zur Überprüfung der Erwartungen hinsichtlich der Periodizität . . . 96
Durchführung der Analyse . . . 96
Zum Verständnis des Periodogramms und der Spektraldichte . . . 98
Zusammenfassung . . . 100
Verwandte Prozeduren . . . 100

Anhänge

A Maße für die Anpassungsgüte *101*

B Ausreißertypen *103*

C Übersicht über ACF-/PACF-Diagramme *104*

D Beispieldateien *108*

Bibliografie *120*

Index *122*

Teil I: Benutzerhandbuch

Kapitel
1

Einführung in Zeitreihen

Eine **Zeitreihe** ist eine Reihe von Beobachtungen, die gewonnen werden, indem eine einzelne Variable regelmäßig über einen Zeitraum beobachtet wird. In einer Zeitreihe mit Bestandsdaten beispielsweise, können die Beobachtungen tägliche Bestandsniveaus für mehrere Monate repräsentieren. Eine Zeitreihe, die den Marktanteil eines Produkts anzeigt, kann aus den wöchentlichen Marktanteilen über den Zeitraum einiger Jahre bestehen. Eine Zeitreihe aus den Gesamtverkaufszahlen kann aus einer Beobachtung pro Monat über den Zeitraum vieler Jahre bestehen. Was alle diese Beispiele gemeinsam haben, ist, dass eine Variable über einen bestimmten Zeitraum in regelmäßigen, bekannten Intervallen beobachtet wurde. Daher ist die Art der Daten für eine typische Zeitreihe eine einzelne Sequenz oder Liste von Beobachtungen, die in regelmäßigen Abständen durchgeführte Messungen repräsentieren.

Tabelle 1-1
Zeitreihe für den täglichen Bestand

Zeit	Woche	Tag	Bestandsniveau
t_1	1	Montag	160
t_2	1	Dienstag	135
t_3	1	Mittwoch	129
t_4	1	Donnerstag	122
t_5	1	Freitag	108
t_6	2	Montag	150
		...	
t_{60}	12	Freitag	120

Eines der wichtigsten Ziele bei der Durchführung einer Zeitreihenanalyse ist der Versuch, die zukünftigen Werte der Zeitreihe vorherzusagen. Ein Modell der Zeitreihe, das die früheren Werte erklärte, kann möglicherweise auch vorhersagen, ob und wie stark die nächsten Werte zu- bzw. abnehmen. Die Fähigkeit, erfolgreich derartige Vorhersagen zu treffen, ist offensichtlich für jeden Betrieb und alle wissenschaftlichen Bereiche von großer Bedeutung.

Zeitreihendaten

Wenn Sie Zeitreihendaten für die Verwendung im Erweiterungsmodul “Trends” definieren, entspricht jede Zeitreihe einer separaten Variablen. Um beispielsweise eine Zeitreihe im Daten-Editor zu definieren, klicken Sie auf die Registerkarte Variablenansicht und geben Sie in eine beliebige leere Zeile einen Variablennamen ein. Jede Beobachtung in einer Zeitreihe entspricht einem Fall (einer Zeile im Daten-Editor).

Wenn Sie eine Tabelle mit Zeitreihendaten öffnen, sollte jede Zeitreihe in einer Spalte der Tabelle angeordnet sein. Wenn Sie bereits eine Tabelle mit Zeitreihen in Zeilen besitzen, können Sie sie dennoch öffnen und mit der Funktion "Transponieren" im Menü "Daten" die Zeilen in Spalten umwandeln.

Transformieren von Daten

Eine Reihe der im Base-System zur Verfügung stehenden Prozeduren für Datentransformationen können bei der Analyse von Zeitreihen nützlich sein.

- Die Prozedur "Datum definieren" (im Menü "Daten") erzeugt Datumsvariablen zum Erstellen von Periodizität und zum Unterscheiden zwischen historischen Perioden, Validierungsperioden und Vorhersageperioden, Trends wurde für die Arbeit mit den durch die Prozedur "Datum definieren" erstellten Daten entwickelt.
- Die Prozedur "Zeitreihen erstellen" (im Menü "Transformieren") erstellt neue Zeitreihenvariablen als Funktionen bestehender Zeitreihenvariablen. Dazu gehören Funktionen, die benachbarte Beobachtungen zum Glätten, zur Durchschnittsberechnung und zur Differenzierung verwenden.
- Die Prozedur "Fehlende Werte ersetzen" (im Menü "Transformieren") ersetzt system- und benutzerdefiniert fehlende Werte durch Schätzwerte auf der Grundlage einer von mehreren möglichen Methoden. Fehlende Daten am Anfang oder Ende einer Zeitreihe stellen kein größeres Problem dar. Sie verkürzen nur die brauchbare Länge der Zeitreihe. Lücken im Inneren einer Zeitreihe (*eingebettete* fehlende Daten) können ein viel schwerwiegenderes Problem darstellen.

Genauere Informationen zu Datentransformationen für Zeitreihen finden Sie im *Benutzerhandbuch für Base*.

Schätz- und Validierungsperioden

Häufig ist es sinnvoll, die Zeitreihe in eine *Schätzperiode* bzw. *historische* Periode und eine *Validierungsperiode* aufzuteilen. Sie entwickeln ein Modell auf der Grundlage der Beobachtungen in der Schätzperiode (historischen Periode) und testen dann in der Validierungsperiode, wie gut es funktioniert. Wenn Sie das Modell zum Vorhersagen von bekannten Punkten (die Punkte in der Validierungsperiode) verwenden, erhalten Sie einen Eindruck von der Güte der Vorhersagen des Modells.

Die Fälle in der Validierungsperiode werden als Prüffälle bezeichnet, da Sie aus dem Erstellungsprozess des Modell ausgeschlossen (und lediglich geprüft) werden. Die Schätzperiode besteht aus den aktuell ausgewählten Fällen in der Arbeitsdatei. Alle verbleibenden Fälle, die auf den letzten ausgewählten Fall folgen, können als Prüffälle verwendet werden. Wenn Sie feststellen, dass das Modell hinreichende Vorhersagen liefert, können Sie die Schätzperiode um die Prüffälle erweitern und anschließend das endgültige Modell erstellen.

Erstellen von Modellen und Vorhersagen

Das Erweiterungsmodul "Trends" bietet zwei Prozeduren, um Modelle zu erstellen und Vorhersagen zu treffen.

- Die Prozedur Zeitreihenmodellierung erstellt Modelle für Zeitreihen und ermöglicht Vorhersagen. Sie enthält den Expert Modeler, der automatisch das beste Modell für jede der Zeitreihen bestimmt. Erfahrene Analysten, die ein größeres Maß an Kontrolle wünschen, finden im Expert Modeler außerdem Werkzeuge zum Erstellen von benutzerdefinierten Modellen.
- Die Prozedur Zeitreihenmodell zuweisen wendet vorhandene, mit der Prozedur "Zeitreihenmodellierung" erstellte Zeitreihenmodelle auf die Arbeitsdatei an. Hierbei können Sie Vorhersagen für Zeitreihen berechnen, für neue oder überarbeitete Daten verfügbar sind, ohne die Modelle erneut zu erstellen. Wenn Sie Grund zu der Annahme haben, dass sich ein Modell verändert hat, können Sie es mit der Prozedur "Zeitreihenmodellierung" erneut erstellen.

Kapitel
2

Zeitreihenmodellierung

Die Prozedur “Zeitreihenmodellierung” schätzt Modelle für die exponentielle Glättung, die univariate ARIMA (Autoregressive Integrated Moving Average) und die multivariate ARIMA (oder Übertragungsfunktionsmodelle) für Zeitreihen und erstellt Prognosen. Die Prozedur enthält den Expert Modeler, der für eine oder mehrere abhängige Variablenreihen automatisch das jeweils am besten angepasste Modell für die ARIMA oder die exponentielle Glättung ermittelt und schätzt, sodass das geeignete Modell nicht mehr nach dem Prinzip von Versuch und Irrtum ermittelt werden muss. Sie können auch ein benutzerdefiniertes ARIMA-Modell oder ein Modell der exponentiellen Glättung angeben.

Beispiel. Als Produktmanager ist es Ihre Aufgabe, für 100 verschiedene Produkte die Verkaufsstückzahlen und Verkaufserlöse für den Folgemonat vorherzusagen. Sie verfügen jedoch nur über geringe oder gar keine Erfahrungen in der Zeitreihenmodellierung. Die historischen Verkaufsstückzahlen für alle 100 Produkten sind in einem einzigen Excel-Arbeitsblatt gespeichert. Nachdem Sie das Arbeitsblatt in SPSS geöffnet haben, fordern Sie im Expert Modeler Vorhersagen für den nächsten Monat an. Der Expert Modeler ermittelt für jedes Produkt das beste Modell für die Verkaufsstückzahlen und erstellt anhand dessen die Vorhersagen. Der Expert Modeler kann mehrere Eingabereihen verarbeiten, sodass Sie die Prozedur nur einmal ausführen müssen, um Vorhersagen für sämtliche Produkte zu erhalten. Wenn Sie die Vorhersagen in der Arbeitsdatei speichern, können Sie die Ergebnisse problemlos zurück nach Excel exportieren.

Statistiken. Maße für die Anpassungsgüte: *R*-Quadrat für stationären Teil, *R*-Quadrat (R^2), Wurzel der mittleren Fehler-Quadratsumme (RMSE), mittlerer absoluter Fehler (MAE), mittlerer absoluter Fehler in Prozent (MAPE), maximaler absoluter Fehler (MaxAE), maximaler absoluter Fehler in Prozent (MaxAPE), normalisiertes Bayes’sches Informationskriterium (BIC). Residuen: Autokorrelationsfunktion, partielle Autokorrelationsfunktion, Ljung-Box-*Q*. Für ARIMA-Modelle: ARIMA-Ordnungen für abhängige Variablen, Übertragungsfunktion-Ordnungen für unabhängige Variablen und Ausreißerschätzer. Außerdem Glättungsparameterschätzer für Modelle mit exponentiellem Glätten.

Diagramme. Diagramme für alle Modelle: Histogramm von *R*-Quadrat für stationären Teil, *R*-Quadrat (R^2), Wurzel der mittleren Fehler-Quadratsumme (RMSE), mittlerer absoluter Fehler (MAE), mittlerer absoluter Fehler in Prozent (MAPE), maximaler absoluter Fehler (MaxAE), maximaler absoluter Fehler in Prozent (MaxAPE), normalisiertes Bayes’sches Informationskriterium (BIC), Boxplots der Residuen-Autokorrelationen und partiellen Autokorrelationen. Ergebnisse für einzelne Modelle: Vorhersagewerte, Anpassungswerte, beobachtete Werte, untere und obere Konfidenzgrenzen, Residuen-Autokorrelationen und partielle Autokorrelationen.

Erläuterungen der Daten für die Prozedur "Zeitreihenmodellierung"

Daten. Die abhängige Variable und etwaige unabhängige Variablen müssen numerisch sein.

Annahmen. Die abhängige Variable und alle unabhängigen Variablen werden als Zeitreihen behandelt, d. h. jeder Fall repräsentiert einen Zeitpunkt, und die nachfolgenden Fälle liegen jeweils ein konstantes Zeitintervall auseinander.

- **Stationarität.** Bei benutzerdefinierten ARIMA-Modellen muss die zu modellierende Zeitreihe stationär sein. Die effektivste Methode zur Transformation einer nichtstationären Zeitreihe in eine stationäre besteht in einer Differenztransformation, aufzurufen über das Dialogfeld "Zeitreihe erstellen".
- **Vorhersagen.** Wenn Vorhersagen anhand von Modellen mit unabhängigen Variablen (Einflussvariablen, Prädiktoren) erstellt werden sollen, muss die Arbeitsdatei für alle Fälle innerhalb der Vorhersageperiode Werte dieser Variablen enthalten. Darüber hinaus dürfen die unabhängigen Variablen in der Schätzperiode keine fehlenden Werte enthalten.

Datum definieren

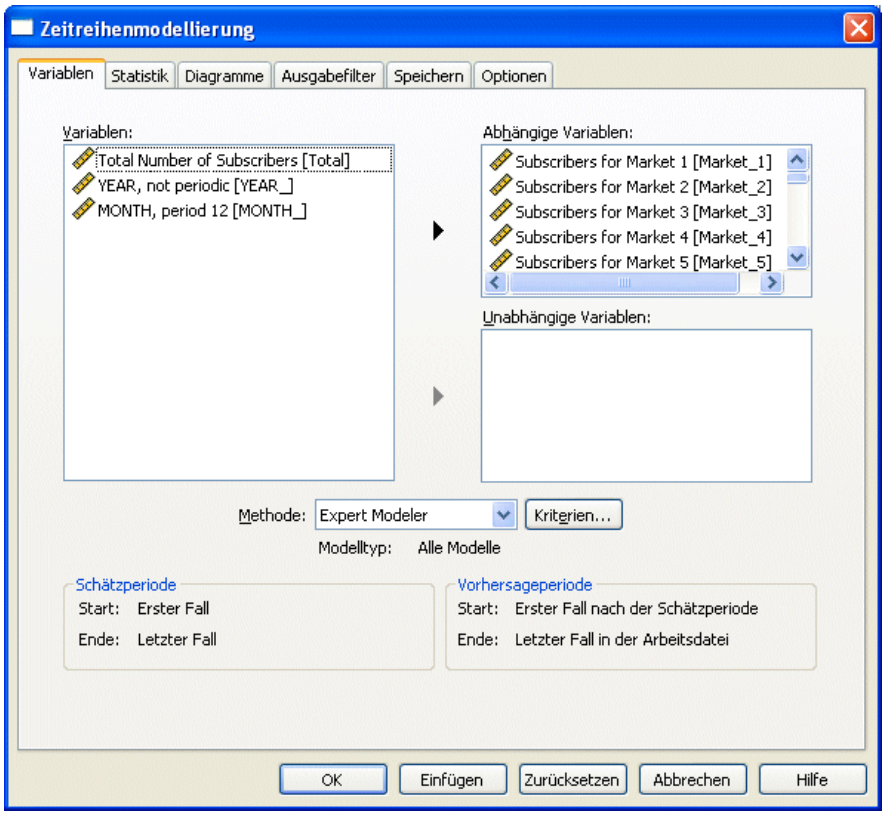
Es ist zwar nicht zwingend erforderlich, aber es empfiehlt sich, im Dialogfeld "Datum definieren" das dem ersten Fall zugewiesene Datum und die Intervalle zwischen den nachfolgenden Fällen anzugeben. Diese Angabe erfolgt vor der Verwendung der Zeitreihenmodellierung und ergibt eine Gruppe von Variablen, die das den einzelnen Fällen zugewiesene Datum bezeichnen. Dabei wird außerdem eine angenommene Periodizität der Daten festgelegt, beispielsweise eine Periodizität von 12, wenn das Zeitintervall zwischen aufeinander folgenden Fällen einen Monat beträgt. Diese Periodizität ist erforderlich, wenn Sie saisonale Modelle erstellen möchten. Wenn Sie keine saisonalen Modelle erstellen möchten und keine Datumsbeschriftungen in der Ausgabe benötigen, müssen Sie das Dialogfeld "Datum definieren" nicht aufrufen. Die den einzelnen Fällen zugewiesene Beschriftung enthält dann einfach die Fallnummer.

So verwenden Sie die Zeitreihenmodellierung:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Modelle erstellen...

Abbildung 2-1
Zeitreihenmodellierung, Registerkarte "Variablen"



- ▶ Wählen Sie auf der Registerkarte "Variablen" mindestens eine abhängige Variable für die Modellierung aus.
- ▶ Wählen Sie in der Dropdown-Liste "Methode" eine Modellierungsmethode aus. Übernehmen Sie für automatische Modellierung die Standardmethode des Expert Modeler. Der Expert Modeler ermittelt dann für jede abhängige Variable das am besten angepasste Modell.

  So erstellen Sie Vorhersagen:

- ▶ Klicken Sie auf die Registerkarte Optionen.
- ▶ Geben Sie die Vorhersageperiode an. Es wird ein Diagramm mit Vorhersagen und beobachteten Werten erstellt.

  Die folgenden Optionen sind verfügbar:

  - Wählen Sie mindestens eine unabhängige Variable aus. Unabhängige Variablen werden weitgehend wie Prädiktoren in der Regressionsanalyse behandelt, sind jedoch optional. Sie können in ARIMA-Modelle, nicht jedoch in Modelle mit exponentiellem Glätten eingeschlossen werden. Wenn Sie Expert Modeler als Modellierungsmethode angeben und unabhängige Variablen einschließen, werden nur ARIMA-Modelle berücksichtigt.
  - Klicken Sie zum Angeben der Modellierungsdetails auf Kriterien.
  - Speichern von Vorhersagen, Konfidenzintervallen und Residuum-Rauschen.

- Speichern der geschätzten Modelle im XML-Format. Gespeicherte Modelle können neuen oder korrigierten Daten zugewiesen werden, um aktualisierte Vorhersagen ohne Neuerstellen der Modelle zu erhalten. Hierfür können Sie die Prozedur Zeitreihenmodell zuweisen verwenden.
- Erstellen von Auswertungsstatistiken für alle geschätzten Modelle.
- Angeben von Übertragungsfunktionen für unabhängige Variablen in ARIMA-Modellen.
- Aktivieren der automatischen Erkennung von Ausreißern.
- Modellspezifische Zeitpunkte als Ausreißer für benutzerdefinierte ARIMA-Modelle.

Modellierungsmethode

Die folgenden Modellierungsmethoden sind verfügbar:

Expert Modeler. Der Expert Modeler ermittelt automatisch das jeweils am besten angepasste Modell für die einzelnen abhängigen Zeitreihen. Wenn unabhängige Variablen (Prädiktoren) angegeben sind, wählt der Expert Modeler für den Einschluss in ARIMA-Modelle diejenigen aus, die eine statistisch signifikante Beziehung mit der abhängigen Zeitreihe aufweisen. Modellvariablen werden ggf. durch Differenzierung und/oder Quadratwurzeltransformation bzw. Transformation mit natürlichem Logarithmus transformiert. Der Expert Modeler berücksichtigt in der Standardeinstellung sowohl Modelle mit exponentiellem Glätten als auch ARIMA-Modelle. Sie können den Expert Modeler jedoch auch auf die ausschließliche Suche nach ARIMA-Modellen bzw. nach Modellen mit exponentiellem Glätten einschränken. Sie können auch die automatische Erkennung von Ausreißern festlegen.

Exponentielles Glätten. Mit dieser Option können Sie ein benutzerdefiniertes Modell mit exponentiellem Glätten angeben. Dabei können Sie zwischen verschiedenen Modellen mit exponentiellem Glätten wählen, die sich hinsichtlich der Behandlung von Trends und Saisonalität unterscheiden.

ARIMA. Mit dieser Option können Sie ein ARIMA-Modell angeben. Hierfür müssen autoregressive Ordnungen, Ordnungen für gleitenden Durchschnitt und der Grad der Differenzierung angegeben werden. Sie können unabhängige Variablen (Prädiktoren) einschließen und für einige oder alle Übertragungsfunktionen definieren. Außerdem können Sie die automatische Erkennung von Ausreißern oder eine bestimmte Gruppe von Ausreißern angeben.

Schätz- und Vorhersageperioden

Schätzperiode. Die Schätzperiode gibt die Gruppe von Fällen an, anhand derer das Modell bestimmt wird. In der Standardeinstellung umfasst die Schätzperiode sämtliche Fälle in der Arbeitsdatei. Sie können die Schätzperiode festlegen, indem Sie im Dialogfeld “Fälle auswählen” die Option Nach Zeit- oder Fallbereich auswählen. Die von der Prozedur verwendete Schätzperiode kann je nach den verfügbaren Daten unterschiedliche abhängige Variablen verwenden und daher vom angezeigten Wert abweichen. Die tatsächliche Schätzperiode für eine bestimmte abhängige Variable entspricht der Zeitspanne, die verbleibt, nachdem die direkt aufeinander folgenden fehlenden Werte von der Variablen entfernt wurden (am Anfang oder am Ende der angegebenen Schätzperiode).

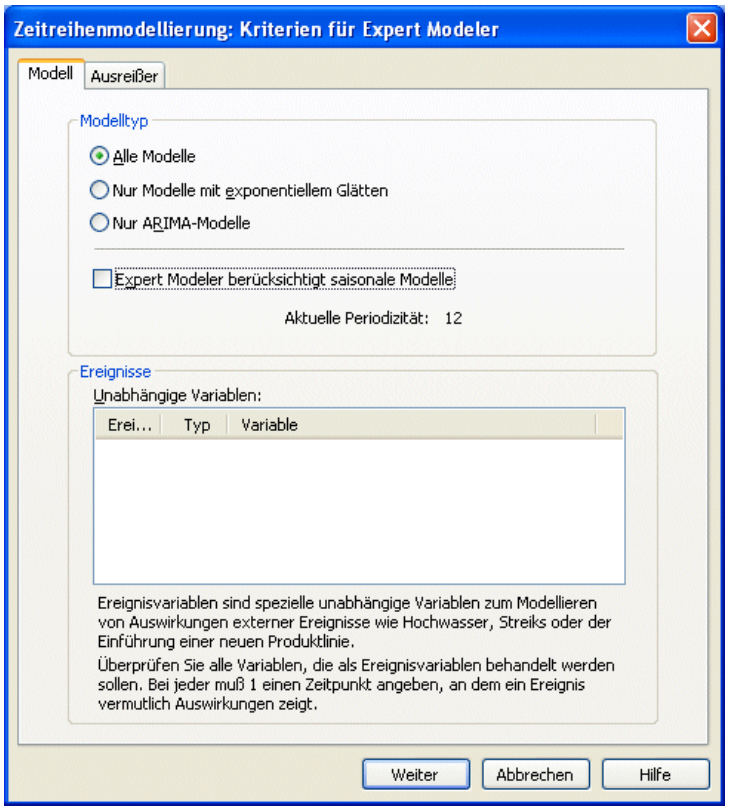
Vorhersageperiode. Die Vorhersageperiode beginnt mit dem ersten Fall nach der Schätzperiode und reicht in der Standardeinstellung bis zum letzten Fall in der Arbeitsdatei. Das Ende der Vorhersageperiode kann auf der Registerkarte Optionen festgelegt werden.

Angeben von Optionen für den Expert Modeler

Der Expert Modeler bietet Optionen, mit denen die Gruppe möglicher Modelle eingeschränkt, die Behandlung von Ausreißern angegeben und Ereignisvariablen angegeben werden können.

Modellauswahl und Ereignisspezifikation

Abbildung 2-2
Dialogfeld "Kriterien für Expert Modeler", Registerkarte "Modell"


Zeitreihenmodellierung: Kriterien für Expert Modeler
Modell
Ausreißer
Modelltyp
Alle Modelle
Nur Modelle mit exponentiellem Glätten
Nur ARIMA-Modelle
Expert Modeler berücksichtigt saisonale Modelle
Aktuelle Periodizität: 12
Ereignisse
Unabhängige Variablen:
Erei...
Typ
Variable
Ereignisvariablen sind spezielle unabhängige Variablen zum Modellieren von Auswirkungen externer Ereignisse wie Hochwasser, Streiks oder der Einführung einer neuen Produktlinie.
Überprüfen Sie alle Variablen, die als Ereignisvariablen behandelt werden sollen. Bei jeder muß 1 einen Zeitpunkt angeben, an dem ein Ereignis vermutlich Auswirkungen zeigt.
Weiter
Abbrechen
Hilfe

Auf der Registerkarte "Modell" können Sie die Modelltypen, die vom Expert Modeler berücksichtigt werden sollen, und Ereignisvariablen angeben.

Modelltyp. Die folgenden Optionen sind verfügbar:

- **Alle Modelle.** Der Expert Modeler berücksichtigt sowohl ARIMA-Modelle als auch Modelle mit exponentiellem Glätten.
- **Nur Modelle mit exponentiellem Glätten.** Der Expert Modeler berücksichtigt nur Modelle mit exponentiellem Glätten.
- **Nur ARIMA-Modelle.** Der Expert Modeler berücksichtigt nur ARIMA-Modelle.

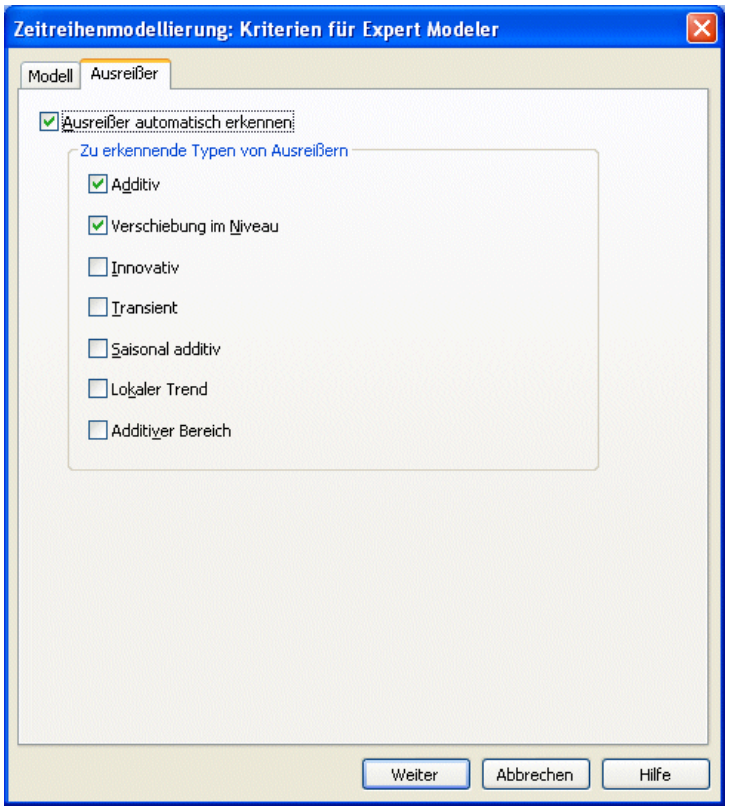
Expert Modeler berücksichtigt saisonale Modelle. Diese Option ist nur verfügbar, wenn für die Arbeitsdatei eine Periodizität definiert wurde. Wenn diese Option aktiviert wird, berücksichtigt der Expert Modeler sowohl saisonale als auch nichtsaisonale Modelle. Wenn diese Option deaktiviert ist, berücksichtigt der Expert Modeler nur nichtsaisonale Modelle.

Aktuelle Periodizität. Gibt die gegenwärtig für die Arbeitsdatei definierte Periodizität an (sofern vorhanden). Die aktuelle Periodizität wird als Ganzzahl angegeben, z. B. 12 für eine jährliche Periodizität, wobei jeder Fall einen Monat darstellt. Wenn keine Periodizität angegeben wurde, wird der Wert *Keine* angezeigt. Für saisonales Modelle muss eine Periodizität angegeben werden. Die Periodizität wird im Dialogfeld "Datum definieren" angegeben.

Ereignisse. Wählen Sie unabhängige Variablen aus, die als Ereignisvariablen behandelt werden sollen. Bei Ereignisvariablen geben Fälle mit einem Wert von 1 die Zeitpunkte an, zu denen erwartet wird, dass die abhängigen Zeitreihen von dem Ereignis beeinflusst werden. Andere Werte als 1 geben an, dass sie nicht beeinflusst werden.

Behandeln von Ausreißern mit dem Expert Modeler

Abbildung 2-3
Dialogfeld "Kriterien für Expert Modeler", Registerkarte "Ausreißer"



Auf der Registerkarte "Ausreißer" können Sie die automatische Erkennung von Ausreißern und die Typen von Ausreißern angeben, die erkannt werden sollen.

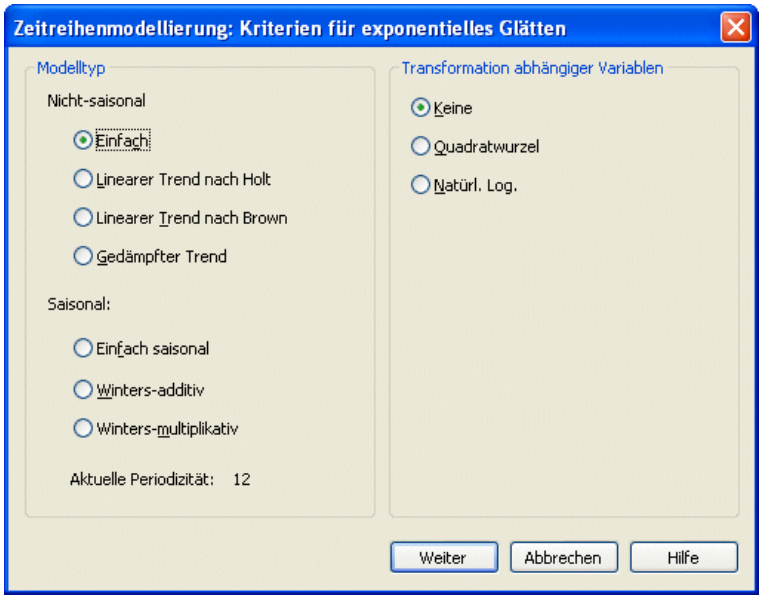
Ausreißer automatisch erkennen. In der Standardeinstellung wird keine automatische Erkennung von Ausreißern durchgeführt. Aktivieren Sie diese Option, um eine automatische Erkennung von Ausreißern durchzuführen, und wählen Sie anschließend mindestens einen der folgenden Ausreißertypen aus:

- Additiv
- Verschiebung im Niveau
- Innovativ
- Transient
- Saisonal additiv
- Lokaler Trend
- Additiver Bereich

Für weitere Informationen siehe Ausreißertypen in Anhang B auf S. 103.

Exponentielles Glätten: Benutzerdefinierte Modelle

Abbildung 2-4
Dialogfeld "Kriterien für exponentielles Glätten"



Modelltyp. Modelle mit exponentiellem Glätten (Gardner, 1985) werden als saisonal oder nichtsaisonal klassifiziert. Saisonale Modelle sind nur verfügbar, wenn für die Arbeitsdatei eine Periodizität definiert wurde (siehe im Folgenden unter "Aktuelle Periodizität").

- **Einfach.** Dieses Modell eignet sich für Zeitreihen, bei denen weder Trend noch Saisonalität vorliegen. Sein einziger Glättungsparameter betrifft das Niveau. Einfaches exponentielles Glätten weist eine sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung null, Differenzenbildung der Ordnung 1, gleitenden Durchschnitten der Ordnung 1 und fehlender Konstante.

- **Holt-Modell mit linearem Trend.** Dieses Modell eignet sich für Zeitreihen, die einen linearen Trend, aber keine Saisonalität aufweisen. Seine Glättungsparameter betreffen Niveau und Trend und es wird angenommen, dass die Werte dieser beiden Elemente unabhängig voneinander sind. Das Holt-Modell ist allgemeiner als das Brown-Modell, aber es kann bei langen Zeitreihen mehr Rechenzeit erfordern. Exponentielles Glätten mit Holt-Modell weist eine sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung 0, Differenzenbildung der Ordnung 2 und gleitenden Durchschnitten der Ordnung 2.
- **Brown-Modell mit linearem Trend.** Dieses Modell eignet sich für Zeitreihen, die einen linearen Trend, aber keine Saisonalität aufweisen. Seine Glättungsparameter betreffen Niveau und Trend und es wird angenommen, dass diese Faktoren gleich sind. Das Brown-Modell ist daher ein Spezialfall des Holt-Modells. Exponentielles Glätten mit dem Brown-Modell weist sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung 0, Differenzenbildung der Ordnung 2 und gleitenden Durchschnitten der Ordnung 2; dabei ist bei den gleitenden Durchschnitten der Koeffizient für den Abstand 2 gleich dem Quadrat der Hälfte des Koeffizienten für den Abstand 1.
- **Gedämpfter Trend.** Dieses Modell eignet sich für Zeitreihen mit einem sich abschwächenden linearen Trend, aber ohne Saisonalität. Seine Glättungsparameter betreffen Niveau, Trend und die Dämpfung des Trends. Exponentielles Glätten mit gedämpftem Trend weist eine sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung 1, Differenzenbildung der Ordnung 1 und gleitenden Durchschnitten der Ordnung 2.
- **Einfach saisonal.** Dieses Modell eignet sich für Zeitreihen ohne Trend, aber mit einem saisonalen Effekt, der über die Zeit gleich bleibt. Seine Glättungsparameter betreffen Niveau und Saison. Einfaches saisonales exponentielles Glätten weist sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung 0, Differenzenbildung der Ordnung 1, saisonaler Differenzenbildung der Ordnung 1 und gleitenden Durchschnitten der Ordnung 1, p und p+1, wobei p die Anzahl der Perioden im Saisonintervall ist (bei monatlichen Daten ist p=12).
- **Winters additiv.** Dieses Modell eignet sich für Zeitreihen mit einem linearen Trend und einem saisonalen Effekt, der nicht vom Niveau der Zeitreihe abhängt. Seine Glättungsparameter betreffen Niveau, Trend und Saison. Exponentielles Glätten mit dem additiven Winters-Modell weist sehr große Ähnlichkeit auf mit einem ARIMA-Modell mit Autoregression der Ordnung 0, Differenzenbildung der Ordnung 1, saisonaler Differenzenbildung der Ordnung 1 und gleitenden Durchschnitten der Ordnung p+1, wobei p die Anzahl der Zeitpunkte in einem Saisonintervall ist (für monatliche Daten ist p=12).
- **Winters multiplikativ.** Dieses Modell eignet sich für Zeitreihen mit einem linearen Trend und einem saisonalen Effekt, der vom Niveau der Zeitreihe abhängt. Seine Glättungsparameter betreffen Niveau, Trend und Saison. Exponentielles Glätten mit dem multiplikativen Winters-Modell weist keine Ähnlichkeit zu irgendeinem ARIMA-Modell auf.

Aktuelle Periodizität. Gibt die gegenwärtig für die Arbeitsdatei definierte Periodizität an (sofern vorhanden). Die aktuelle Periodizität wird als Ganzzahl angegeben, z. B. 12 für eine jährliche Periodizität, wobei jeder Fall einen Monat darstellt. Wenn keine Periodizität angegeben wurde, wird der Wert *Keine* angezeigt. Für saisonales Modelle muss eine Periodizität angegeben werden. Die Periodizität wird im Dialogfeld “Datum definieren” angegeben.

Transformation abhängiger Variablen. Sie können angeben, dass für alle abhängigen Variablen vor der Modellierung eine Transformation durchgeführt wird.

- **Keine.** Es wird keine Transformation durchgeführt.
- **Quadratwurzel.** Quadratwurzeltransformation.
- **Natürlicher Logarithmus.** Transformation mit natürlichem Logarithmus.

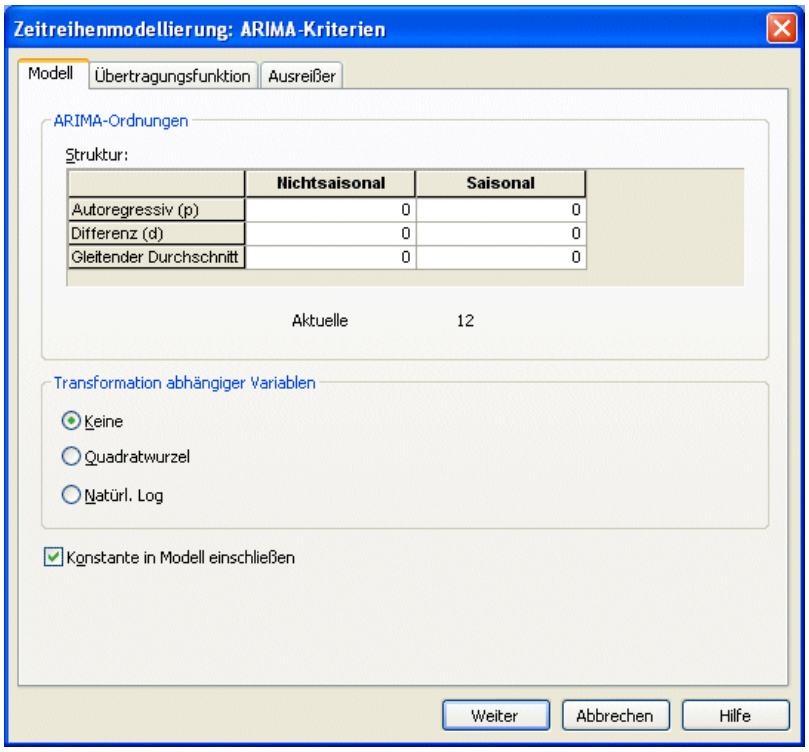
Benutzerdefinierte ARIMA-Modelle

Mit der Zeitreihenmodellierung können benutzerdefinierte nichtsaisonale oder saisonale univariate ARIMA-Modelle (auch als Box-Jenkins-Modelle (Box, Jenkins, als auch Reinsel, 1994) bekannt) mit oder ohne feste Gruppe von Einflussvariablen (Prädiktoren) erstellt werden. (ARIMA steht für Autoregressive Integrated Moving Average), engl. für autoregressiver integrierter gleitender Durchschnitt. Sie können für einige oder alle Prädiktoren Übertragungsfunktionen definieren, die automatische Erkennung von Ausreißern festlegen oder eine bestimmte Gruppe von Ausreißern angeben.

- Alle auf der Registerkarte "Variablen" angegebenen unabhängigen Variablen (Prädiktoren) werden explizit in das Modell eingeschlossen. Im Gegensatz dazu werden im Expert Modeler unabhängige Variablen nur eingeschlossen, wenn sie eine statistisch signifikante Beziehung mit der abhängigen Variablen aufweisen.

Modellspezifikation für benutzerdefinierte ARIMA-Modelle

Abbildung 2-5
Dialogfeld "ARIMA-Kriterien", Registerkarte "Modell"



Auf der Registerkarte “Modelle” können Sie die Struktur eines benutzerdefinierten ARIMA-Modells festlegen.

ARIMA-Ordnungen. Geben Sie Werte für die verschiedenen ARIMA-Komponenten des Modells in die entsprechenden Zellen des Strukturgitters ein. Alle Werte müssen nicht-negative Ganzzahlen sein. Bei autoregressiven Komponenten und Komponenten des gleitenden Durchschnitts stellt der Wert die höchste Ordnung dar. Alle positiven niedrigeren Ordnungen werden in das Modell eingeschlossen. Wenn Sie beispielsweise 2 angeben, enthält das Modell die Ordnungen 2 und 1. Die Zellen in der Spalte “Saisonal” sind nur verfügbar, wenn für die Arbeitsdatei eine Periodizität definiert wurde (siehe im Folgenden unter “Aktuelle Periodizität”).

- **Autoregressiv (p).** Die Anzahl autoregressiver Ordnungen im Modell. Autoregressive Ordnungen geben die zurückliegenden Werte der Zeitreihe an, die für die Vorhersage der aktuellen Werte verwendet werden. Eine autoregressive Ordnung von 2 gibt beispielsweise an, dass die Werte der Zeitreihe, die zwei Zeitperioden zurückliegt, für die Vorhersage der aktuellen Werte verwendet wird.
- **Differenz (d).** Gibt die Ordnung der Differenzierung an, die vor dem Schätzen der Modelle auf die Zeitreihe angewendet wurde. Differenzierung ist erforderlich, wenn Trends vorhanden sind. (Zeitreihen mit Trends sind normalerweise nichtstationär, und bei der ARIMA-Modellierung wird Stationarität angenommen.) Mithilfe der Differenzierung werden die Effekte der Trends entfernt. Die Ordnung der Differenzierung entspricht dem Grad des Trends der Zeitreihe: Differenzierung erster Ordnung erklärt lineare Trends, Differenzierung zweiter Ordnung erklärt quadratische Trends usw.
- **Gleitender Durchschnitt (q).** Die Anzahl von Ordnungen des gleitenden Durchschnitts im Modell. Ordnungen des gleitenden Durchschnitts geben an, wie Abweichungen vom Mittelwert der Zeitreihe für zurückliegende Werte zum Vorhersagen der aktuellen Werte verwendet werden. Ordnungen des gleitenden Durchschnitts von 1 und 2 geben beispielsweise an, dass beim Vorhersagen der aktuellen Werte der Zeitreihe Abweichungen vom Mittelwert der Zeitreihe von den beiden letzten Zeitperioden berücksichtigt werden sollen.

Saisonale Ordnungen. Saisonale autoregressive Komponenten, Komponenten des gleitenden Durchschnitts und Differenzierungskomponenten entsprechen im Prinzip ihren nichtsaisonalen Gegenstücken. Bei saisonalen Ordnungen werden die Werte der aktuellen Zeitreihe jedoch von Werten zurückliegender Zeitreihen beeinflusst, die um eine oder mehrere saisonalen Perioden getrennt sind. Bei monatlichen Daten (saisonale Periode von 12) beispielsweise bedeutet eine saisonale Ordnung von 1, dass der Wert der aktuellen Zeitreihe durch den Zeitreihenwert beeinflusst wird, der 12 Perioden vor dem aktuellen liegt. Eine saisonale Ordnung von 1 entspricht bei monatlichen Daten einer nichtsaisonalen Ordnung von 12.

Aktuelle Periodizität. Gibt die gegenwärtig für die Arbeitsdatei definierte Periodizität an (sofern vorhanden). Die aktuelle Periodizität wird als Ganzzahl angegeben, z. B. 12 für eine jährliche Periodizität, wobei jeder Fall einen Monat darstellt. Wenn keine Periodizität angegeben wurde, wird der Wert *Keine* angezeigt. Für saisonales Modelle muss eine Periodizität angegeben werden. Die Periodizität wird im Dialogfeld “Datum definieren” angegeben.

Transformation abhängiger Variablen. Sie können angeben, dass für alle abhängigen Variablen vor der Modellierung eine Transformation durchgeführt wird.

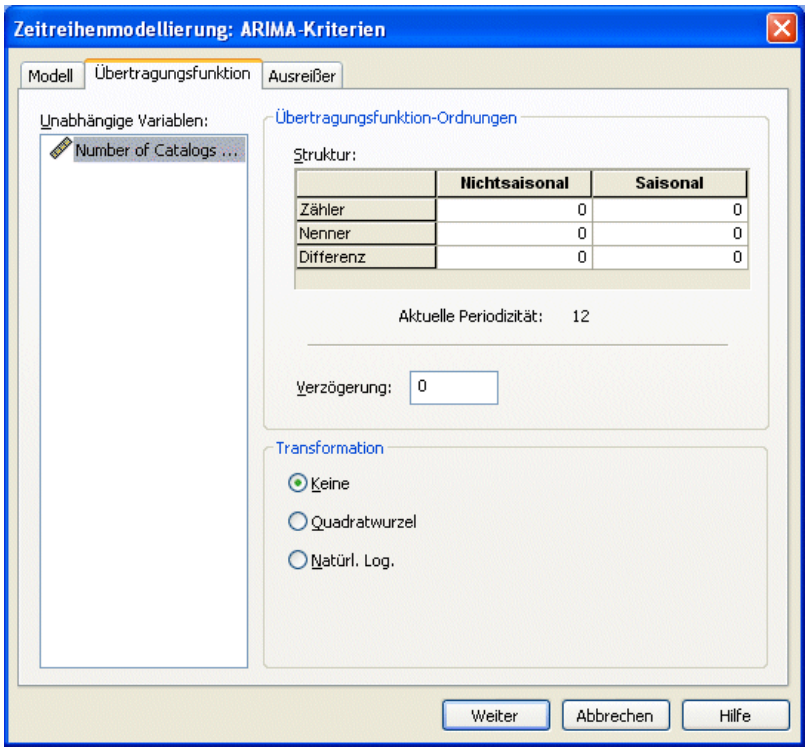
- **Keine.** Es wird keine Transformation durchgeführt.

- **Quadratwurzel.** Quadratwurzeltransformation.
- **Natürlicher Logarithmus.** Transformation mit natürlichem Logarithmus.

Konstante in Modell einschließen. Der Einschluss einer Konstanten ist das Standardverfahren, sofern Sie nicht sicher wissen, dass der Gesamtmittelwert der Zeitreihe 0 ist. Bei der Anwendung von Differenzierung empfiehlt es sich, die Konstante auszuschließen.

Übertragungsfunktionen in benutzerdefinierten ARIMA-Modellen

Abbildung 2-6
Dialogfeld "ARIMA-Kriterien", Registerkarte "Übertragungsfunktion"



Auf der Registerkarte "Übertragungsfunktion" (nur vorhanden, wenn unabhängige Variablen angegeben wurden) können Sie Übertragungsfunktionen für bestimmte oder alle unabhängigen Variablen definieren, die auf der Registerkarte "Variablen" angegeben sind. Durch Übertragungsfunktionen kann angegeben werden, wie zurückliegende Werte unabhängiger Variablen (Prädiktoren) zum Vorhersagen zukünftiger Werte der abhängigen Zeitreihe verwendet werden sollen.

Übertragungsfunktion-Ordnungen. Geben Sie Werte für die verschiedenen Komponenten der Übertragungsfunktion in die entsprechenden Zellen des Strukturgitters ein. Alle Werte müssen nichtnegative Ganzzahlen sein. Bei Zähler- und Nennerkomponenten stellt der Wert die höchste Ordnung dar. Alle positiven niedrigeren Ordnungen werden in das Modell eingeschlossen. Darüber hinaus wird die Ordnung 0 bei Zählerkomponenten immer eingeschlossen. Wenn Sie beispielsweise 2 als Zähler angeben, enthält das Modell die Ordnungen 2, 1 und 0. Wenn Sie 3 als Nenner angeben, enthält das Modell die Ordnungen 3, 2 und 1. Die Zellen in der Spalte

"Saisonal" sind nur verfügbar, wenn für die Arbeitsdatei eine Periodizität definiert wurde (siehe im Folgenden unter "Aktuelle Periodizität").

- **Zähler.** Der Zähler-Term der Übertragungsfunktion. Dieser gibt an, welche zurückliegenden Werte aus der ausgewählten unabhängigen Zeitreihe (Prädiktoren) zum Vorhersagen der aktuellen Werte der abhängigen Zeitreihe verwendet werden. Ein Zähler-Term von 1 gibt beispielsweise an, dass der Wert einer unabhängigen Zeitreihe, die eine Periode zurückliegt, und der aktuelle Wert der unabhängigen Zeitreihe zum Vorhersagen des aktuellen Werts der einzelnen abhängigen Zeitreihen verwendet werden.
- **Nenner.** Der Nenner-Term der Übertragungsfunktion. Dieser gibt an, wie Abweichungen vom Mittelwert der Zeitreihe für zurückliegende Werte der ausgewählten unabhängigen Zeitreihe (Prädiktoren) zum Vorhersagen der aktuellen Werte der abhängigen Zeitreihe verwendet werden. Ein Nenner-Term von 1 gibt beispielsweise an, dass beim Vorhersagen der aktuellen Werte für die einzelnen abhängigen Zeitreihen Abweichungen vom Mittelwert einer unabhängigen Zeitreihe berücksichtigt werden sollen, die eine Zeitperiode zurückliegt.
- **Differenz.** Gibt die Ordnung der Differenzierung an, die vor dem Schätzen der Modelle auf die ausgewählte unabhängige Zeitreihe (Prädiktoren) angewendet wurde. Wenn Trends vorhanden sind, ist die Differenzierung erforderlich, um die Effekte der Trends zu entfernen.

Saisonale Ordnungen. Saisonale Zähler-, Nenner- und Differenzierungskomponenten entsprechen im Prinzip ihren nichtsaisonalen Gegenstücken. Bei saisonalen Ordnungen werden die Werte der aktuellen Zeitreihe jedoch von Werten zurückliegender Zeitreihen beeinflusst, die um eine oder mehrere saisonalen Perioden getrennt sind. Bei monatlichen Daten (saisonale Periode von 12) beispielsweise bedeutet eine saisonale Ordnung von 1, dass der Wert der aktuellen Zeitreihe durch den Zeitreihenwert beeinflusst wird, der 12 Perioden vor dem aktuellen liegt. Eine saisonale Ordnung von 1 entspricht bei monatlichen Daten einer nichtsaisonalen Ordnung von 12.

Aktuelle Periodizität. Gibt die gegenwärtig für die Arbeitsdatei definierte Periodizität an (sofern vorhanden). Die aktuelle Periodizität wird als Ganzzahl angegeben, z. B. 12 für eine jährliche Periodizität, wobei jeder Fall einen Monat darstellt. Wenn keine Periodizität angegeben wurde, wird der Wert *Keine* angezeigt. Für saisonales Modelle muss eine Periodizität angegeben werden. Die Periodizität wird im Dialogfeld "Datum definieren" angegeben.

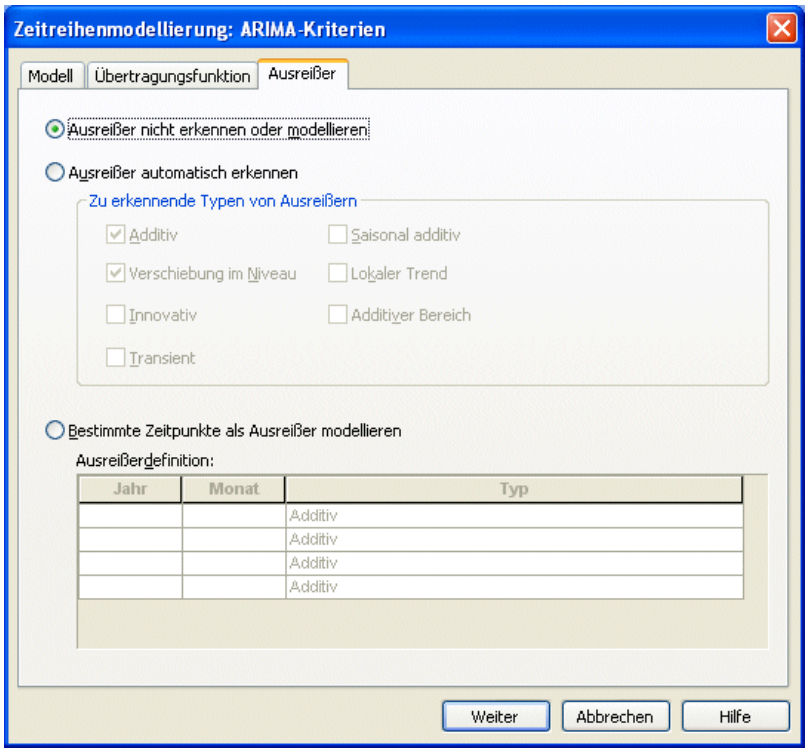
Verzögerung. Wenn eine Verzögerung festgelegt wird, verzögert sich der Einfluss der unabhängigen Variable um die Anzahl der angegebenen Intervalle. Bei einer Verzögerung mit dem Wert 5 beeinflusst der Wert der unabhängigen Variable zum Zeitpunkt t die Vorhersagen erst nach dem Ablauf von fünf Perioden ($t + 5$).

Transformation. Die Angabe einer Übertragungsfunktion für ein Set von unabhängigen Variablen enthält auch eine optionale Transformation, die für diese Variablen ausgeführt werden soll.

- **Keine.** Es wird keine Transformation durchgeführt.
- **Quadratwurzel.** Quadratwurzeltransformation.
- **Natürlicher Logarithmus.** Transformation mit natürlichem Logarithmus.

Ausreißer in benutzerdefinierten ARIMA-Modellen

Abbildung 2-7
Dialogfeld "ARIMA-Kriterien", Registerkarte "Ausreißer"



Auf der Registerkarte "Ausreißer" sind die folgenden Möglichkeiten für die Behandlung von Ausreißern verfügbar (Pena, Tiao, als auch Tsay, 2001): Ausreißer automatisch erkennen, bestimmte Punkte als Ausreißer angeben oder Ausreißer nicht erkennen bzw. modellieren.

Ausreißer nicht erkennen oder modellieren. In der Standardeinstellung werden Ausreißer weder erkannt noch modelliert. Aktivieren Sie diese Option, um die Erkennung und Modellierung von Ausreißern zu deaktivieren.

Ausreißer automatisch erkennen. Aktivieren Sie diese Option, um eine automatische Erkennung von Ausreißern durchzuführen, und wählen Sie mindestens einen der folgenden Ausreißertypen aus:

- Additiv
- Verschiebung im Niveau
- Innovativ
- Transient
- Saisonal additiv
- Lokaler Trend
- Additiver Bereich

Für weitere Informationen siehe Ausreißertypen in Anhang B auf S. 103.

Bestimmte Zeitpunkte als Ausreißer modellieren. Aktivieren Sie diese Option, um bestimmte Zeitpunkte als Ausreißer zu modellieren. Verwenden Sie für jeden Ausreißer eine eigene Zeile im Ausreißerdefinitionsgitter. Geben Sie in alle Zellen einer bestimmten Zeile Werte ein.

- **Typ.** Der Ausreißertyp. Folgende Typen werden unterstützt: additiv (Standardeinstellung), Verschiebung im Niveau, innovativ, transient, saisonal additiv und lokaler Trend.

Anmerkung 1: Wenn für die Arbeitsdatei kein Datum angegeben wurde, wird im Ausreißerdefinitionsgitter nur die Spalte *Beobachtung* angezeigt. Um eine Ausreißer festzulegen, geben Sie die Zeilennummer des entsprechenden Falles ein (wie im Daten-Editor angezeigt).

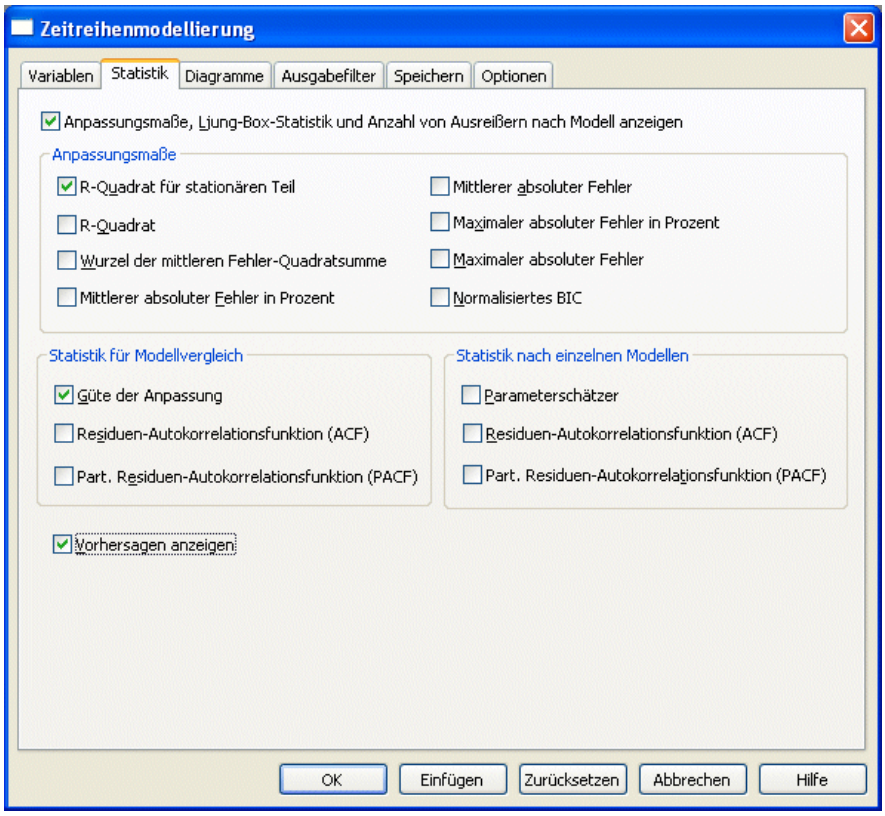
Anmerkung 2: Die Spalte *Zyklus* im Ausreißerdefinitionsgitter (sofern vorhanden) bezieht sich auf die Werte der Variablen *CYCLE_* in der Arbeitsdatei.

Ausgabe

Als Ausgabe stehen sowohl Ergebnisse für einzelne Modelle als auch für alle Modelle errechnete Ergebnisse zur Verfügung. Die Ergebnisse für einzelne Modelle können über vom Benutzer angegebene Kriterien auf eine Gruppe von am besten/schlechtesten angepassten Modellen eingeschränkt werden.

Statistiken und Vorhersagetabellen

Abbildung 2-8
Zeitreihenmodellierung, Registerkarte "Statistik"



Die Registerkarte “Statistik” bietet Optionen für die Anzeige von Tabellen mit den Modellierungsergebnissen.

Anpassungsmaße, Ljung-Box-Statistik und Anzahl von Ausreißern nach Modell anzeigen. Aktivieren Sie diese Option, um für jedes geschätzte Modell eine Tabelle mit den ausgewählten Anpassungsmaße, Ljung-Box-Werten und der Anzahl von Ausreißern anzuzeigen.

Anpassungsmaße. Sie können eine oder mehrere der folgenden Optionen für die Aufnahme in die Tabelle auswählen, die die Anpassungsmaße für die einzelnen geschätzten Modelle enthält:

- *R*-Quadrat für stationären Teil
- *R*-Quadrat
- Wurzel der mittleren Fehler-Quadratsumme
- Mittlerer absoluter Fehler in Prozent
- Mittlerer absoluter Fehler
- Maximaler absoluter Fehler in Prozent
- Maximaler absoluter Fehler
- Normalisiertes BIC

Für weitere Informationen siehe Maße für die Anpassungsgüte in Anhang A auf S. 101.

Statistik für Modellvergleich. Mit dieser Gruppe von Optionen wird die Anzeige der Tabellen gesteuert, die die für alle Modelle berechneten Statistiken enthalten. Für jede Option wird eine eigene Tabelle erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- **Güte der Anpassung.** Tabelle der Auswertungsstatistiken und Perzentile für *R*-Quadrat für stationären Teil, *R*-Quadrat, Wurzel der mittleren Fehler-Quadratsumme, mittlerer absoluter Fehler in Prozent, mittlerer absoluter Fehler, maximaler absoluter Fehler in Prozent, maximaler absoluter Fehler und normalisiertes Bayes-Informationskriterium.
- **Residuen-Autokorrelationsfunktion (ACF).** Tabelle der Auswertungsstatistiken und Perzentile für Autokorrelationen der Residuen für alle geschätzten Modelle.
- **Part. Residuen-Autokorrelationsfunktion (PACF).** Tabelle der Auswertungsstatistiken und Perzentile für partielle Autokorrelationen der Residuen für alle geschätzten Modelle.

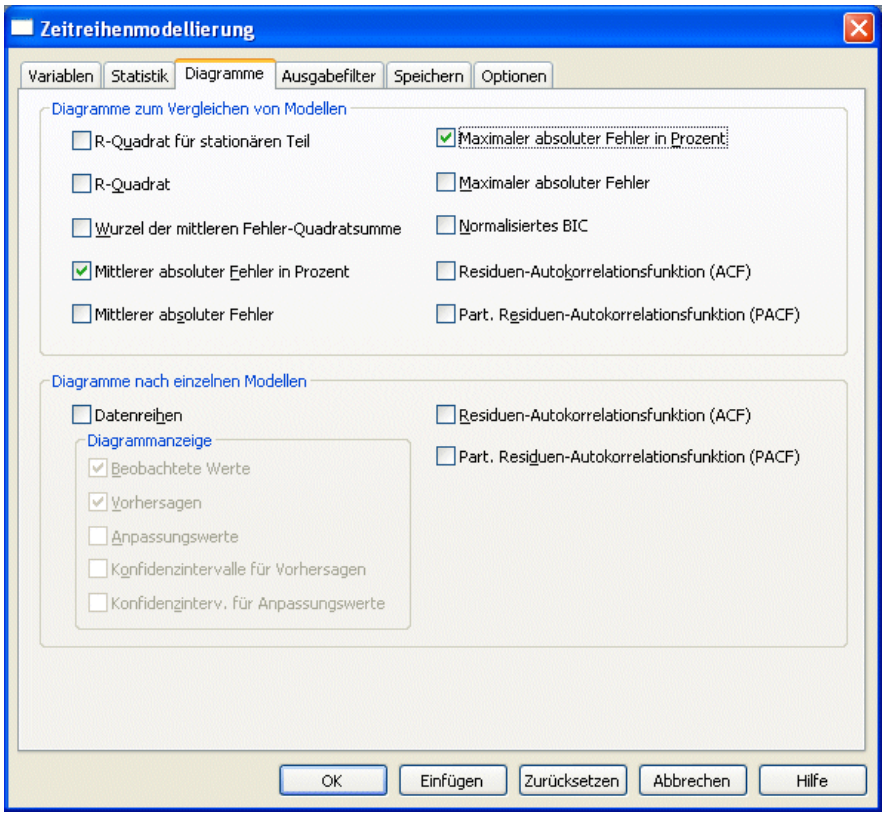
Statistik nach einzelnen Modellen. Mit dieser Gruppe von Optionen wird die Anzeige der Tabellen gesteuert, die detaillierte Informationen zu den einzelnen geschätzten Modellen enthalten. Für jede Option wird eine eigene Tabelle erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- **Parameterschätzer.** Zeigt für jedes geschätzte Modell eine Tabelle der Parameterschätzer an. Für Modelle mit exponentiellem Glätten und ARIMA-Modelle werden eigene Tabellen angezeigt. Wenn Ausreißer vorhanden sind, werden Parameterschätzer für diese ebenfalls in einer eigenen Tabelle angezeigt.
- **Residuen-Autokorrelationsfunktion (ACF).** Zeigt für jedes geschätzte Modell eine Tabelle der Residuen-Autokorrelationen nach Intervall an. Die Tabelle enthält die Konfidenzintervalle für die Autokorrelationen.
- **Part. Residuen-Autokorrelationsfunktion (PACF).** Zeigt für jedes geschätzte Modell eine Tabelle der partiellen Residuen-Autokorrelationen nach Intervall an. Die Tabelle enthält die Konfidenzintervalle für die partiellen Autokorrelationen.

Vorhersagen anzeigen. Zeigt eine Tabelle der Modellvorhersagen und der Konfidenzintervalle für jedes geschätzte Modell an. Die Vorhersageperiode wird auf der Registerkarte “Optionen” festgelegt.

Diagramme

Abbildung 2-9
Zeitreihenmodellierung, Registerkarte “Diagramme”



Die Registerkarte “Diagramme” bietet Optionen für die Anzeige von Diagrammen der Modellierungsergebnisse.

Diagramme zum Vergleichen von Modellen

Mit dieser Gruppe von Optionen wird die Anzeige der Diagramme gesteuert, die die für alle Modelle berechneten Statistiken enthalten. Für jede Option wird ein eigenes Diagramm erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- *R*-Quadrat für stationären Teil
- *R*-Quadrat
- Wurzel der mittleren Fehler-Quadratsumme
- Mittlerer absoluter Fehler in Prozent
- Mittlerer absoluter Fehler

- Maximaler absoluter Fehler in Prozent
- Maximaler absoluter Fehler
- Normalisiertes BIC
- Residuen-Autokorrelationsfunktion (ACF)
- Part. Residuen-Autokorrelationsfunktion (PACF)

Für weitere Informationen siehe Maße für die Anpassungsgüte in Anhang A auf S. 101.

Diagramme nach einzelnen Modellen

Datenreihen. Aktivieren Sie diese Option, um für die einzelnen geschätzten Modelle Diagramme der vorhergesagten Werte zu erhalten. Sie können eine oder mehrere der folgenden Optionen für die Aufnahme in das Diagramm auswählen:

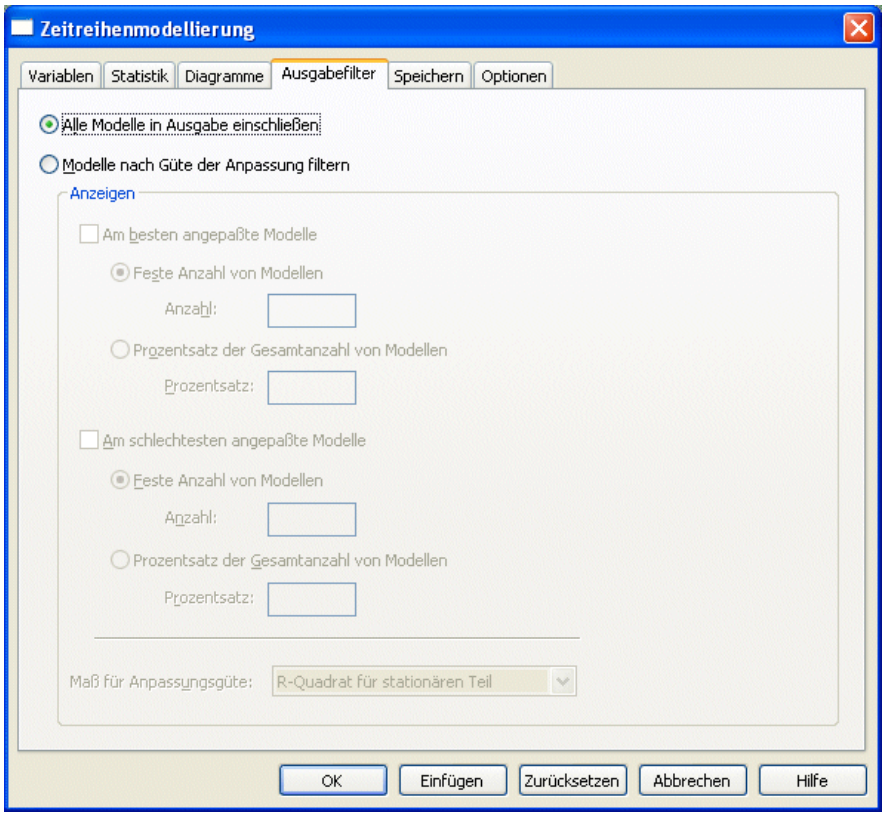
- **Beobachtete Werte.** Die beobachteten Werte der abhängigen Zeitreihe.
- **Vorhersagen.** Die vom Modell vorhergesagten Werte für die Vorhersageperiode.
- **Anpassungswerte.** Die vom Modell vorhergesagten Werte für die Schätzperiode.
- **Konfidenzintervalle für Vorhersagen.** Die Konfidenzintervalle für die Vorhersageperiode.
- **Konfidenzintervalle für Anpassungswerte.** Die Konfidenzintervalle für die Schätzperiode.

Residuen-Autokorrelationsfunktion (ACF). Zeigt für jedes geschätzte Modell ein Diagramm der Residuen-Autokorrelationen an.

Part. Residuen-Autokorrelationsfunktion (PACF). Zeigt für jedes geschätzte Modell ein Diagramm der partiellen Residuen-Autokorrelationen an.

Einschränken der Ausgabe auf die am besten/schlechtesten angepassten Modelle

Abbildung 2-10
Zeitreihenmodellierung, Registerkarte "Ausgabefilter"



Die Registerkarte "Ausgabefilter" enthält Optionen, mit denen sowohl Tabellen- als auch Diagrammausgabe auf eine Untergruppe von geschätzten Modellen beschränkt werden kann. Sie können die Ausgabe mithilfe von Anpassungskriterien auf die am besten und/oder am schlechtesten angepassten Modelle beschränken. In der Standardeinstellung sind alle geschätzten Modelle in der Ausgabe enthalten.

Am besten angepaßte Modelle. Aktivieren Sie diese Option, um die am besten angepassten Modelle in die Ausgabe einzuschließen. Wählen Sie ein Maß für Anpassungsgüte und geben Sie die Anzahl von Modellen an, die eingeschlossen werden sollen. Die Auswahl dieser Option hindert Sie nicht daran, auch die am schlechtesten angepassten Modelle auszuwählen. Die Ausgabe würde in diesem Fall sowohl die am schlechtesten als auch die am besten angepassten Modelle enthalten.

- **Feste Anzahl von Modellen.** Gibt an, dass Ergebnisse für die *n* am besten angepassten Modelle angezeigt werden. Wenn die Anzahl die Anzahl der geschätzten Modelle überschreitet, werden alle Modelle angezeigt.
- **Prozentsatz der Gesamtanzahl von Modellen.** Gibt an, dass Ergebnisse für Modelle angezeigt werden, deren Werte für die Anpassungsgüte für alle geschätzten Modelle in den obersten *n* Prozent liegen.

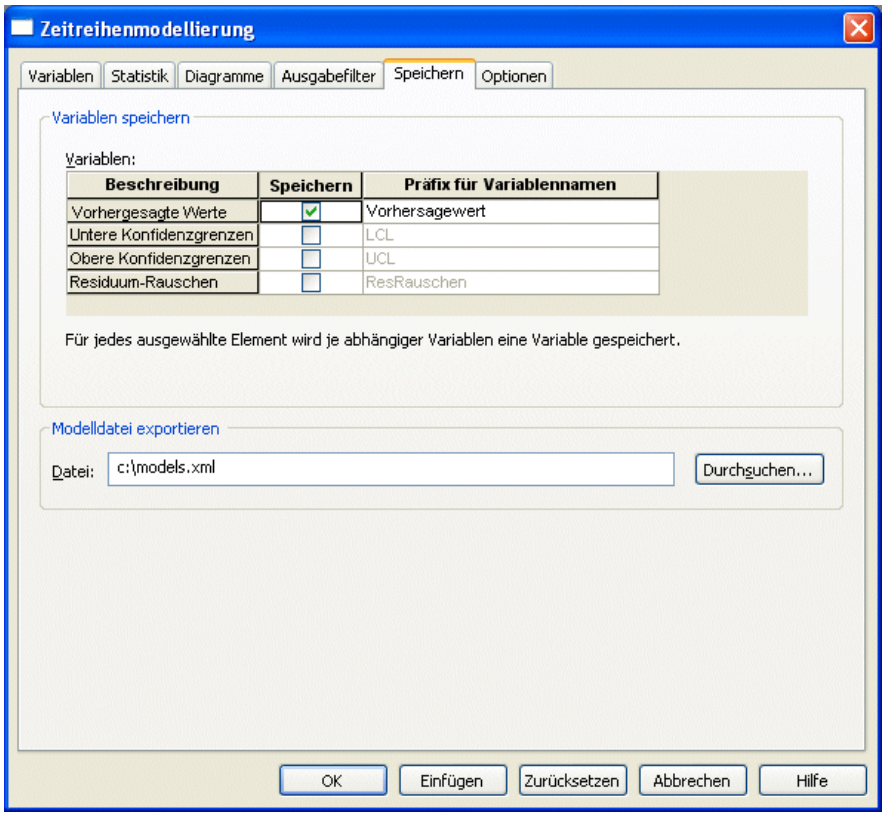
Am schlechtesten angepaßte Modelle. Aktivieren Sie diese Option, um die am schlechtesten angepassten Modelle in die Ausgabe einzuschließen. Wählen Sie ein Maß für Anpassungsgüte und geben Sie die Anzahl von Modellen an, die eingeschlossen werden sollen. Die Auswahl dieser Option hindert Sie nicht daran, auch die am besten angepassten Modelle auszuwählen. Die Ausgabe würde in diesem Fall sowohl die am besten als auch die am schlechtesten angepassten Modelle enthalten.

- **Feste Anzahl von Modellen.** Gibt an, dass Ergebnisse für die *n* am schlechtesten angepassten Modelle angezeigt werden. Wenn die Anzahl die Anzahl der geschätzten Modelle überschreitet, werden alle Modelle angezeigt.
- **Prozentsatz der Gesamtanzahl von Modellen.** Gibt an, dass Ergebnisse für Modelle angezeigt werden, deren Werte für die Anpassungsgüte für alle geschätzten Modelle in den untersten *n* Prozent liegen.

Maß für Anpassungsgüte. Wählen Sie das Maß für die Anpassungsgüte, anhand dessen die Modelle gefiltert werden sollen. Der Standardwert ist *R*-Quadrat für den stationären Teil.

Speichern von Modellvorhersagen und Modellspezifikationen

Abbildung 2-11
Zeitreihenmodellierung, Registerkarte "Speichern"



Auf der Registerkarte “Speichern” können Sie festlegen, dass Modellvorhersagen als neue Variablen in der Arbeitsdatei gespeichert werden sollen. Außerdem können die Modellspezifikationen in einer externen Datei im XML-Format gespeichert werden.

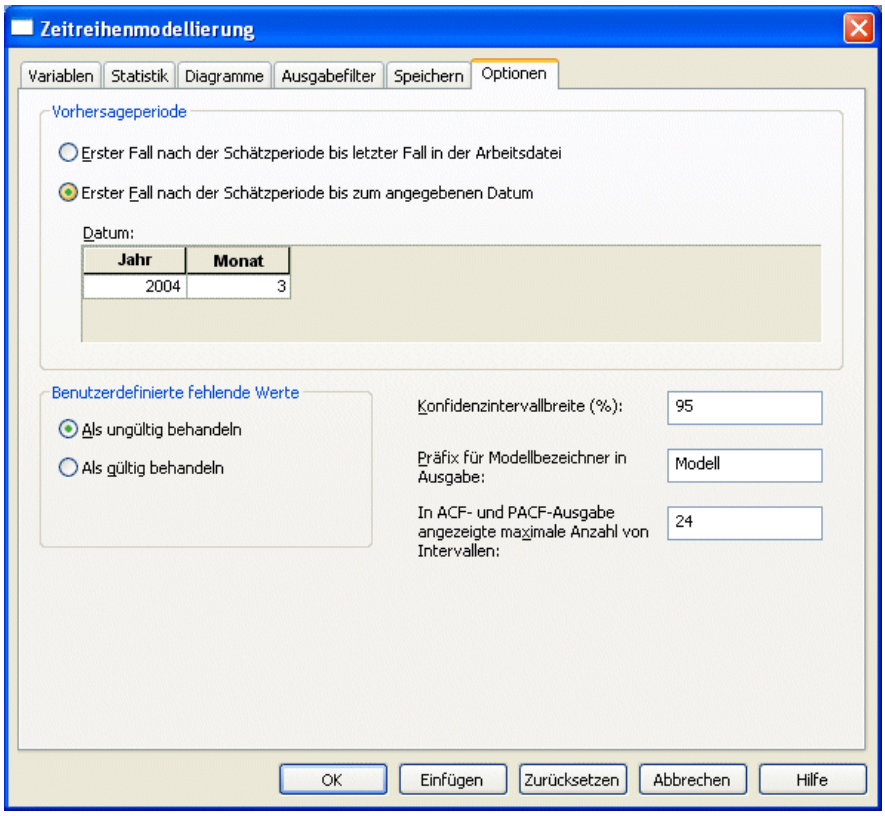
Variablen speichern. Sie können Modellvorhersagen, Konfidenzintervalle und Residuen als neue Variablen in der Arbeitsdatei speichern. Jede abhängige Zeitreihe erzeugt eine eigene Gruppe neuer Variablen, und jede neue Variable enthält Werte für Schätz- und Vorhersageperioden. Wenn die Vorhersageperiode die Länge der abhängigen Zeitreihe überschreitet, werden neue Fälle hinzugefügt. Wenn die neuen Variablen gespeichert werden sollen, aktivieren Sie jeweils das zugehörige Kontrollkästchen “Speichern”. In der Standardeinstellung werden keine neuen Variablen gespeichert.

- **Vorhergesagte Werte.** Die vom Modell vorhergesagten Werte.
- **Untere Konfidenzgrenzen.** Die unteren Konfidenzgrenzen für die vorhergesagten Werte.
- **Obere Konfidenzgrenzen.** Die oberen Konfidenzgrenzen für die vorhergesagten Werte.
- **Residuum-Rauschen.** Die Modellresiduen. Bei der Transformation von abhängigen Variablen (z. B mit natürlichem Logarithmus) sind dies die Residuen für die transformierte Reihe.
- **Präfix für Variablennamen.** Geben Sie Präfixe für die Namen der neuen Variablen an oder übernehmen Sie die Standardpräfixe. Variablennamen bestehen aus dem Präfix, dem Namen der zugewiesenen abhängigen Variablen und einem Modellbezeichner. Beim Auftreten von Namenskonflikten wird der Variablenname ggf. erweitert. Das Präfix muss den Regeln für gültige Variablennamen entsprechen.

Modelldatei exportieren. Modellspezifikationen für alle geschätzten Modelle werden in die angegebene XML-Datei exportiert. Mit gespeicherten Modellen können anhand von aktuelleren Daten aktualisierte Vorhersagen erstellt werden. Verwenden Sie hierfür die Prozedur Zeitreihenmodell zuweisen.

Optionen

Abbildung 2-12
Zeitreihenmodellierung, Registerkarte "Optionen"



Auf der Registerkarte "Optionen" können Sie die Vorhersageperiode, die Behandlung fehlender Werte und die Breite des Konfidenzintervalls festlegen, ein benutzerdefiniertes Präfix für Modellbezeichner angeben und die für die Autokorrelationen angezeigte Anzahl von Intervallen festlegen.

Vorhersageperiode. Die Vorhersageperiode beginnt immer mit dem ersten Fall nach dem Ende der Schätzperiode (d. h. der Gruppe von Fällen, die das Modell bestimmt) und endet entweder mit dem letzten Fall in der Arbeitsdatei oder an einem vom Benutzer festgelegten Datum. In der Standardeinstellung endet die Schätzperiode mit dem letzten Fall in der Arbeitsdatei. Das Ende kann jedoch im Dialogfeld "Fälle auswählen" durch Auswahl der Option Nach Zeit- oder Fallbereich geändert werden.

- **Erster Fall nach der Schätzperiode bis letzter Fall in der Arbeitsdatei.** Wählen Sie diese Option aus, wenn das Ende der Schätzperiode vor dem letzten Fall in der Arbeitsdatei liegt und die Vorhersage auch den letzten Fall erfassen soll. Über diese Option werden üblicherweise

Vorhersagen für eine Prüfperiode erstellt, sodass die Modellvorhersagen mit einer Teilmenge der tatsächlichen Werte verglichen werden können.

- **Erster Fall nach der Schätzperiode bis zum angegebenen Datum.** Wählen Sie diese Option aus, um das Ende der Vorhersageperiode explizit festzulegen. Über diese Option werden üblicherweise Vorhersagen erstellt, die über das Ende der tatsächlichen Zeitreihe hinausreichen. Geben Sie in alle Zellen des Datengitters Werte ein.

  Wenn für die Arbeitsdatei kein Datum angegeben wurde, wird im Datengitter nur die Spalte *Beobachtung* angezeigt. Um das Ende der Vorhersageperiode festzulegen, geben Sie die Zeilennummer des entsprechenden Falles ein (wie im Daten-Editor angezeigt).

  Die Spalte *Zyklus* im Datumsgitter (sofern vorhanden) bezieht sich auf die Werte der Variablen *CYCLE_* in der Arbeitsdatei.

Benutzerdefinierte fehlende Werte. Mit diesen Optionen wird die Behandlung benutzerdefinierter fehlender Werte gesteuert.

- **Als ungültig behandeln.** Benutzerdefiniert fehlende Werte werden wie systemdefiniert fehlende Werte behandelt.
- **Als gültig behandeln.** Benutzerdefinierte fehlende Werte werden als gültige Werte behandelt.

Richtlinie für fehlende Werte. Die folgenden Regeln gelten bei der Modellierungsprozedur für die Behandlung fehlender Werte. Sie gelten auch für systemdefinierte fehlende Wert und als ungültig behandelte benutzerdefinierte fehlende Werte:

- Fälle mit fehlenden Werten einer abhängigen Variablen, die innerhalb der Schätzperiode liegen, werden in das Modell eingeschlossen. Die genaue Behandlung des fehlenden Werts hängt von der Schätzmethode ab.
- Wenn eine unabhängige Variable innerhalb der Schätzperiode fehlende Werte aufweist, wird eine Warnung ausgegeben. Im Expert Modeler werden Modelle, die die unabhängige Variable enthalten, ohne die Variable geschätzt. Bei einer benutzerdefinierten ARIMA werden Modelle, die die unabhängige Variable enthalten, nicht geschätzt.
- Wenn unabhängige Variablen innerhalb der Vorhersageperiode fehlende Werte aufweisen, gibt die Prozedur eine Warnung aus und führt die Vorhersage auf der Grundlage der vorhandenen Werte aus.

Konfidenzintervallbreite (%). Konfidenzintervalle werden für die Modellvorhersagen und Residuen-Autokorrelationen berechnet. Es kann ein beliebiger positiver Wert unter 100 angegeben werden. In der Standardeinstellung wird ein Konfidenzintervall von 95 % verwendet.

Präfix für Modellbezeichner in Ausgabe. Jede auf der Registerkarte “Variablen” angegebene abhängige Variable erzeugt ein eigenes geschätztes Modell. Jedes Modell weist einen eindeutigen Namen auf, der sich aus einem anpassbarem Präfix und einem ganzzahligen Suffix zusammensetzt. Sie können ein Präfix eingeben oder das vorgegebene Präfix unter *Model* übernehmen.

In ACF- und PACF-Ausgabe angezeigte maximale Anzahl von Intervallen. Sie können die Höchstanzahl von Intervallen festlegen, die in Tabellen und Diagrammen für Autokorrelationen und partielle Autokorrelationen angezeigt werden.

Zusätzliche Funktionen beim Befehl TSMODEL

Sie können die Zeitreihenmodellierung an Ihre Bedürfnisse anpassen, indem Sie ihre Auswahl in ein Syntax-Fenster einfügen und die resultierende Befehlssyntax für den Befehl `TSMODEL` bearbeiten. Mit der Befehlssyntax-Sprache verfügen Sie über folgende Möglichkeiten:

- Angeben der saisonalen Periode für die Daten (mit dem Schlüsselwort `SEASONLENGTH` im Unterbefehl `AUXILIARY`). Dadurch wird die aktuelle Periodizität (sofern vorhanden) überschrieben, die bis dahin in der Arbeitsdatei gültig war.
- Angeben von nicht aufeinander folgenden Intervallen für Komponenten der benutzerdefinierten ARIMA und der Übertragungsfunktion (mit den Unterbefehlen `ARIMA` und `TRANSFERFUNCTION`). Sie können beispielsweise ein benutzerdefiniertes ARIMA-Modell mit autoregressiven Intervallen der Ordnungen 1, 3 und 6 oder eine Übertragungsfunktion mit Zählerintervallen der Ordnungen 2, 5 und 8 angeben.
- Angeben mehrerer Gruppen von Modellierungsspezifikationen (z. B. Modellierungsmethode, ARIMA-Ordnungen, unabhängige Variablen usw.) für eine Ausführung der Prozedur "Zeitreihenmodellierung" (mit dem Unterbefehl `MODEL`).

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
3

Zeitreihenmodell zuweisen

Mit der Prozedur “Zeitreihenmodell zuweisen” werden vorhandene Zeitreihenmodelle aus einer externen Datei geladen und auf die Arbeitsdatei angewendet. Mit dieser Prozedur können Sie Vorhersagen für Zeitreihen berechnen, für die neue oder überarbeitete Daten verfügbar sind, ohne die Modelle erneut zu erstellen. Modelle werden mit der Prozedur Zeitreihenmodellierung erzeugt.

Beispiel. Sie arbeiten als Bestandsmanager für einen Großhandel und sind für 5,000 Produkte zuständig. Mit dem Expert Modeler haben Sie Modelle erstellt, die den Absatz der einzelnen Produkte innerhalb der nächsten drei Monate vorhersagen. Das Data Warehouse wird jeden Monat mit den neuen Verkaufsdaten aktualisiert, mit denen Sie Ihre Vorhersagen monatlich aktualisieren möchten. Verwenden Sie hierfür die Prozedur “Zeitreihenmodell zuweisen”, bei der Sie die ursprünglichen Modellen beibehalten können und lediglich die Modellparameter neu schätzen müssen, um die neuen Daten zu berücksichtigen.

Statistiken. Maße für die Anpassungsgüte: *R*-Quadrat für stationären Teil, *R*-Quadrat (R^2), Wurzel der mittleren Fehler-Quadratsumme (RMSE), mittlerer absoluter Fehler (MAE), mittlerer absoluter Fehler in Prozent (MAPE), maximaler absoluter Fehler (MaxAE), maximaler absoluter Fehler in Prozent (MaxAPE), normalisiertes Bayes’sches Informationskriterium (BIC). Residuen: Autokorrelationsfunktion, partielle Autokorrelationsfunktion, Ljung-Box-*Q*.

Diagramme. Diagramme für alle Modelle: Histogramm von *R*-Quadrat für stationären Teil, *R*-Quadrat (R^2), Wurzel der mittleren Fehler-Quadratsumme (RMSE), mittlerer absoluter Fehler (MAE), mittlerer absoluter Fehler in Prozent (MAPE), maximaler absoluter Fehler (MaxAE), maximaler absoluter Fehler in Prozent (MaxAPE), normalisiertes Bayes’sches Informationskriterium (BIC), Boxplots der Residuen-Autokorrelationen und partiellen Autokorrelationen. Ergebnisse für einzelne Modelle: Vorhersagewerte, Anpassungswerte, beobachtete Werte, untere und obere Konfidenzgrenzen, Residuen-Autokorrelationen und partielle Autokorrelationen.

Erläuterungen der Daten für die Prozedur “Zeitreihenmodell zuweisen”

Daten. Die Variablen (abhängige und unabhängige), denen Modelle zugewiesen werden, müssen numerisch sein.

Annahmen. Modelle werden Variablen in der Arbeitsdatei zugewiesen, die die gleichen Namen wie die im Modell angegebenen Variablen tragen. Alle diese Variablen werden als Zeitreihen behandelt, d. h. jeder Fall repräsentiert einen Zeitpunkt, und die nachfolgenden Fälle liegen jeweils ein konstantes Zeitintervall auseinander.

- **Vorhersagen.** Wenn Vorhersagen anhand von Modellen mit unabhängigen Variablen (Einflussvariablen, Prädiktoren) erstellt werden sollen, muss die Arbeitsdatei für alle Fälle innerhalb der Vorhersageperiode Werte dieser Variablen enthalten. Wenn die Modellparameter

erneut geschätzt werden, dürfen die unabhängigen Variablen in der Schätzperiode keine fehlenden Werte enthalten.

Datum definieren

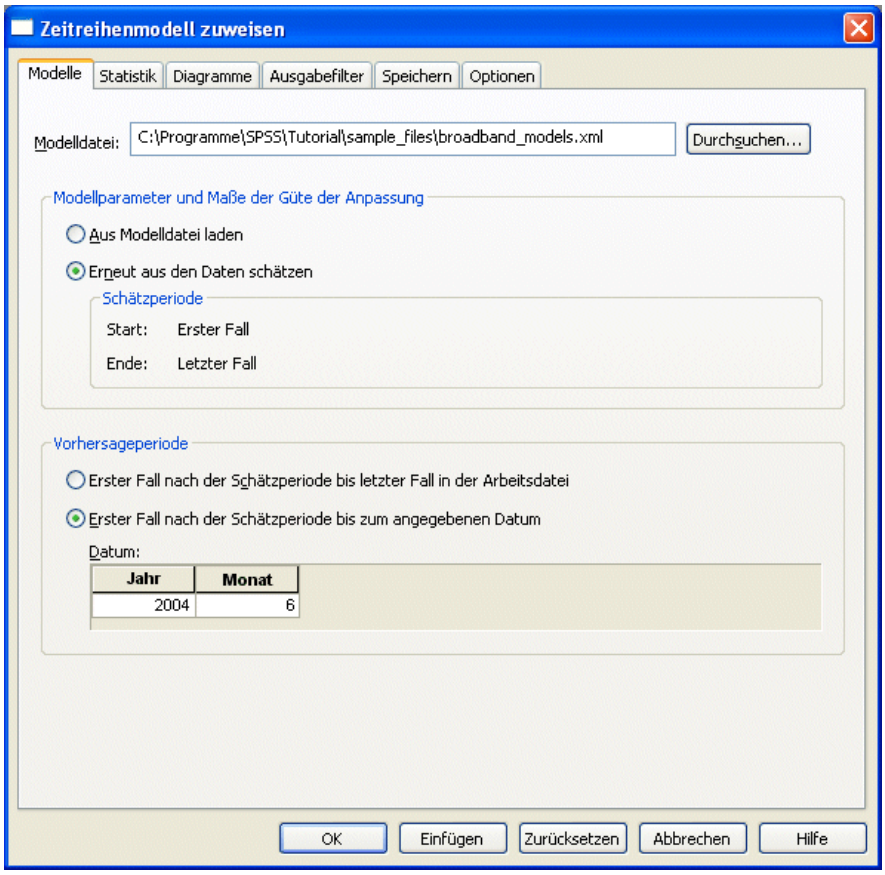
Für die Prozedur “Zeitreihenmodell zuweisen” muss die Periodizität (sofern vorhanden) der Arbeitsdatei mit der Periodizität der zuzuweisenden Modelle übereinstimmen. Wenn Sie bei der Vorhersage das gleiche Daten-Set (möglicherweise mit neuen oder korrigierten Daten) verwenden, mit dem auch das Modell erstellt wurde, ist diese Bedingung erfüllt. Wenn für die Arbeitsdatei keine Periodizität vorliegt, können Sie zum Dialogfeld “Datum definieren” wechseln und dort eine Periodizität erstellen. Wurden die Modelle hingegen ohne Angabe einer Periodizität erstellt, darf die Arbeitsdatei auch keine Periodizität aufweisen.

So weisen Sie Modelle zu:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Modelle zuweisen...

Abbildung 3-1
Zeitreihenmodell zuweisen, Registerkarte “Modelle”



▶ Geben Sie die Angaben für eine Modelldatei ein oder klicken Sie auf Durchsuchen und wählen Sie eine Modelldatei aus (Modelldateien werden mit der Prozedur Zeitreihenmodellierung erstellt).

Die folgenden Optionen sind verfügbar:

- Erneutes Schätzen der Modellparameter mithilfe der Daten in der Arbeitsdatei. Vorhersagen werden anhand der erneut geschätzten Parameter erstellt.
- Speichern von Vorhersagen, Konfidenzintervallen und Residuum-Rauschen.
- Speichern des erneut geschätzten Modells im XML-Format.

Modellparameter und Maße für die Anpassungsgüte

Laden aus der Modelldatei. Vorhersagen werden mithilfe der Modellparameter aus der Modelldatei erstellt, wobei die Parameter nicht erneut geschätzt werden. Die Maße für die Anpassungsgüte, die in der Ausgabe angezeigt und zum Filtern von Modellen (am besten/schlechtesten angepasste Modelle) verwendet werden, werden aus der Modelldatei abgerufen und entsprechen den Daten, die bei der Entwicklung (oder letzten Aktualisierung) des jeweiligen Modells verwendet wurden. Bei dieser Option werden bei Vorhersagen keine historischen Daten (für entweder abhängige oder unabhängige Variablen) in der Arbeitsdatei berücksichtigt. Wählen Sie Erneut aus den Daten schätzen aus, wenn historische Daten in die Vorhersage einbezogen werden sollen. Darüber hinaus werden bei Vorhersagen keine Werte der abhängigen Zeitreihe in der Vorhersageperiode einbezogen, Werte unabhängiger Variablen werden hingegen berücksichtigt. Wenn Sie aktuellere Werte der abhängigen Zeitreihe haben und diese in die Vorhersagen einbeziehen möchten, müssen Sie eine erneute Schätzung vornehmen und die Schätzperiode entsprechend korrigieren.

Erneut aus den Daten schätzen. Modellparameter werden mithilfe der Daten in der Arbeitsdatei erneut geschätzt. Eine erneute Schätzung der Modellparameter wirkt sich nicht auf die Modellstruktur aus. Ein ARIMA(1,0,1)-Modell bleibt beispielsweise unverändert, die autoregressiven Parameter und die Parameter für gleitenden Durchschnitt werden hingegen erneut geschätzt. Bei der erneuten Schätzung werden keine neuen Ausreißer erkannt. Ausreißer werden, sofern vorhanden, immer aus der Modelldatei abgerufen.

- **Schätzperiode.** Die Schätzperiode gibt die Gruppe von Fällen an, anhand derer die Modellparameter erneut geschätzt werden. In der Standardeinstellung umfasst die Schätzperiode sämtliche Fälle in der Arbeitsdatei. Sie können die Schätzperiode festlegen, indem Sie im Dialogfeld “Fälle auswählen” die Option Nach Zeit- oder Fallbereich auswählen. Die von der Prozedur verwendete Schätzperiode kann je nach den verfügbaren Daten unterschiedliche Modelle verwenden und daher vom angezeigten Wert abweichen. Die tatsächliche Schätzperiode für ein bestimmtes Modell entspricht der Zeitspanne, die verbleibt, nachdem die direkt aufeinander folgenden fehlenden Werte von der abhängigen Variablen des Modells entfernt wurden (am Anfang oder am Ende der angegebenen Schätzperiode).

Vorhersageperiode

Die Vorhersageperiode für die einzelnen Modelle beginnt immer mit dem ersten Fall nach dem Ende der Schätzperiode und endet entweder mit dem letzten Fall in der Arbeitsdatei oder an einem vom Benutzer festgelegten Datum. Wenn die Parameter nicht erneut geschätzt werden (Standardeinstellung), ist die Schätzperiode für die einzelnen Modelle die Gruppe von Fällen, die bei der Entwicklung (oder letzten Aktualisierung) des jeweiligen Modells verwendet wurden.

- **Erster Fall nach der Schätzperiode bis letzter Fall in der Arbeitsdatei.** Wählen Sie diese Option aus, wenn das Ende der Schätzperiode vor dem letzten Fall in der Arbeitsdatei liegt und die Vorhersage auch den letzten Fall erfassen soll.
- **Erster Fall nach der Schätzperiode bis zum angegebenen Datum.** Wählen Sie diese Option aus, um das Ende der Vorhersageperiode explizit festzulegen. Geben Sie in alle Zellen des Datengitters Werte ein.

  Wenn für die Arbeitsdatei kein Datum angegeben wurde, wird im Datengitter nur die Spalte *Beobachtung* angezeigt. Um das Ende der Vorhersageperiode festzulegen, geben Sie die Zeilennummer des entsprechenden Falles ein (wie im Daten-Editor angezeigt).

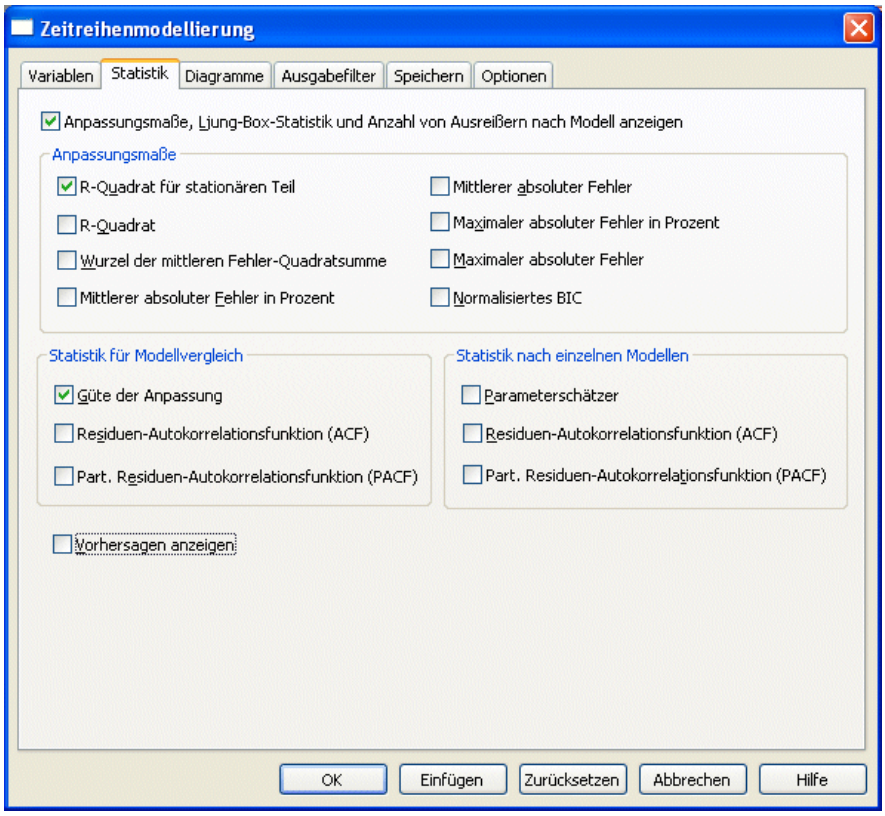
  Die Spalte *Zyklus* im Datumsgitter (sofern vorhanden) bezieht sich auf die Werte der Variablen *CYCLE_* in der Arbeitsdatei.

Ausgabe

Als Ausgabe stehen sowohl Ergebnisse für einzelne Modelle als auch Ergebnisse für alle Modelle zur Verfügung. Die Ergebnisse für einzelne Modelle können über vom Benutzer angegebene Kriterien auf eine Gruppe von am besten/schlechtesten angepassten Modellen eingeschränkt werden.

Statistiken und Vorhersagetabellen

Abbildung 3-2
Zeitreihenmodell zuweisen, Registerkarte "Statistik"



Die Registerkarte "Statistik" enthält Optionen, mit denen Tabellen für Statistiken zur Güte der Anpassung, Modellparameter, Autokorrelationsfunktionen und Vorhersagen angezeigt werden können. Sofern die Modellparameter nicht erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte "Modelle"), stammen die angezeigten Werte für die Anpassungsmaße, Ljung-Box-Werte und Modellparameter aus der Modelldatei und entsprechen den Daten, die bei der Entwicklung (oder letzten Aktualisierung) des jeweiligen Modells verwendet wurden. Informationen zu Ausreißern werden immer aus der Modelldatei abgerufen.

Anpassungsmaße, Ljung-Box-Statistik und Anzahl von Ausreißern nach Modell anzeigen. Aktivieren Sie diese Option, um für jedes Modell eine Tabelle mit den ausgewählten Anpassungsmaße, Ljung-Box-Werten und der Anzahl von Ausreißern anzuzeigen.

Anpassungsmaße. Sie können eine oder mehrere der folgenden Optionen für die Aufnahme in die Tabelle auswählen, die die Anpassungsmaße für die einzelnen Modelle enthält:

- *R*-Quadrat für stationären Teil
- *R*-Quadrat
- Wurzel der mittleren Fehler-Quadratsumme
- Mittlerer absoluter Fehler in Prozent

- Mittlerer absoluter Fehler
- Maximaler absoluter Fehler in Prozent
- Maximaler absoluter Fehler
- Normalisiertes BIC

Für weitere Informationen siehe Maße für die Anpassungsgüte in Anhang A auf S. 101.

Statistik für Modellvergleich. Mit dieser Gruppe von Optionen wird die Anzeige der Tabellen gesteuert, die die Statistiken für alle Modelle enthalten. Für jede Option wird eine eigene Tabelle erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- **Güte der Anpassung.** Tabelle der Auswertungsstatistiken und Perzentile für *R*-Quadrat für stationären Teil, *R*-Quadrat, Wurzel der mittleren Fehler-Quadratsumme, mittlerer absoluter Fehler in Prozent, mittlerer absoluter Fehler, maximaler absoluter Fehler in Prozent, maximaler absoluter Fehler und normalisiertes Bayes-Informationskriterium.
- **Residuen-Autokorrelationsfunktion (ACF).** Tabelle der Auswertungsstatistiken und Perzentile für Autokorrelationen der Residuen für alle geschätzten Modelle. Diese Tabelle ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).
- **Part. Residuen-Autokorrelationsfunktion (PACF).** Tabelle der Auswertungsstatistiken und Perzentile für partielle Autokorrelationen der Residuen für alle geschätzten Modelle. Diese Tabelle ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).

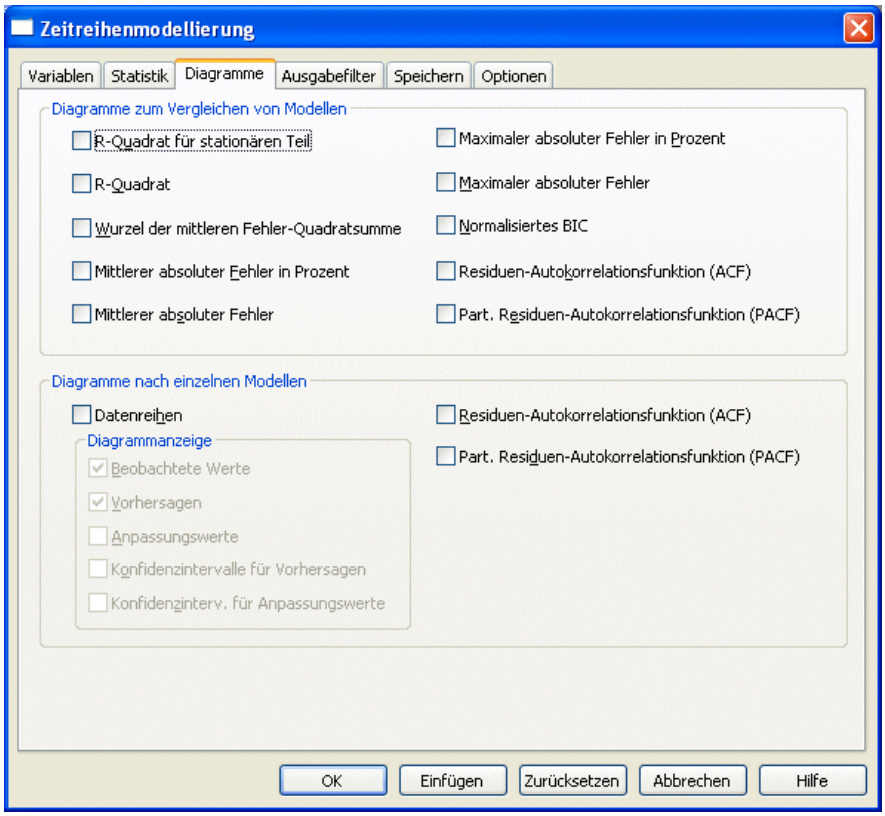
Statistik nach einzelnen Modellen. Mit dieser Gruppe von Optionen wird die Anzeige der Tabellen gesteuert, die detaillierte Informationen zu den einzelnen Modellen enthalten. Für jede Option wird eine eigene Tabelle erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- **Parameterschätzer.** Zeigt für jedes Modell eine Tabelle der Parameterschätzer an. Für Modelle mit exponentiellem Glätten und ARIMA-Modelle werden eigene Tabellen angezeigt. Wenn Ausreißer vorhanden sind, werden Parameterschätzer für diese ebenfalls in einer eigenen Tabelle angezeigt.
- **Residuen-Autokorrelationsfunktion (ACF).** Zeigt für jedes geschätzte Modell eine Tabelle der Residuen-Autokorrelationen nach Intervall an. Die Tabelle enthält die Konfidenzintervalle für die Autokorrelationen. Diese Tabelle ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).
- **Part. Residuen-Autokorrelationsfunktion (PACF).** Zeigt für jedes geschätzte Modell eine Tabelle der partiellen Residuen-Autokorrelationen nach Intervall an. Die Tabelle enthält die Konfidenzintervalle für die partiellen Autokorrelationen. Diese Tabelle ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).

Vorhersagen anzeigen. Zeigt eine Tabelle der Modellvorhersagen und der Konfidenzintervalle für jedes Modell an.

Diagramme

Abbildung 3-3
Zeitreihenmodell zuweisen, Registerkarte "Diagramme"



Die Registerkarte "Diagramme" enthält Optionen, mit denen Diagramme für Statistiken zur Güte der Anpassung, Autokorrelationsfunktionen und Zeitreihenwerte (einschließlich Vorhersagen) angezeigt werden können.

Diagramme zum Vergleichen von Modellen

Mit dieser Gruppe von Optionen wird die Anzeige der Diagramme gesteuert, die die Statistiken für alle Modelle enthalten. Sofern die Modellparameter nicht erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte "Modelle"), stammen die angezeigten Werte aus der Modelldatei und entsprechen den Daten, die bei der Entwicklung (oder letzten Aktualisierung) des jeweiligen Modells verwendet wurden. Autokorrelationsdiagramme sind zudem nur verfügbar, wenn die Modellparameter erneut geschätzt werden. Für jede Option wird ein eigenes Diagramm erstellt. Sie können eine oder mehrere der folgenden Optionen auswählen:

- *R*-Quadrat für stationären Teil
- *R*-Quadrat
- Wurzel der mittleren Fehler-Quadratsumme
- Mittlerer absoluter Fehler in Prozent
- Mittlerer absoluter Fehler

- Maximaler absoluter Fehler in Prozent
- Maximaler absoluter Fehler
- Normalisiertes BIC
- Residuen-Autokorrelationsfunktion (ACF)
- Part. Residuen-Autokorrelationsfunktion (PACF)

Für weitere Informationen siehe Maße für die Anpassungsgüte in Anhang A auf S. 101.

Diagramme nach einzelnen Modellen

Datenreihen. Aktivieren Sie diese Option, um für die einzelnen Modelle Diagramme der vorhergesagten Werte zu erhalten. Beobachtete Werte, Anpassungswerte, Konfidenzintervalle für Anpassungswerte und Autokorrelationen sind nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte "Modelle"). Sie können eine oder mehrere der folgenden Optionen für die Aufnahme in das Diagramm auswählen:

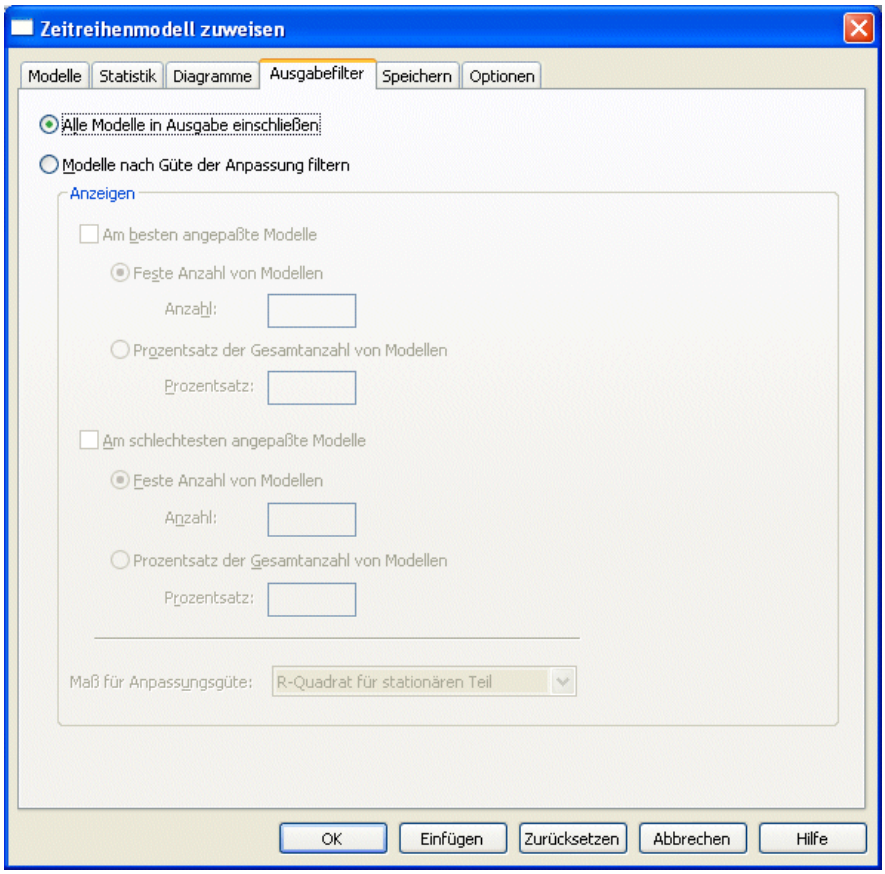
- **Beobachtete Werte.** Die beobachteten Werte der abhängigen Zeitreihe.
- **Vorhersagen.** Die vom Modell vorhergesagten Werte für die Vorhersageperiode.
- **Anpassungswerte.** Die vom Modell vorhergesagten Werte für die Schätzperiode.
- **Konfidenzintervalle für Vorhersagen.** Die Konfidenzintervalle für die Vorhersageperiode.
- **Konfidenzintervalle für Anpassungswerte.** Die Konfidenzintervalle für die Schätzperiode.

Residuen-Autokorrelationsfunktion (ACF). Zeigt für jedes geschätzte Modell ein Diagramm der Residuen-Autokorrelationen an.

Part. Residuen-Autokorrelationsfunktion (PACF). Zeigt für jedes geschätzte Modell ein Diagramm der partiellen Residuen-Autokorrelationen an.

Einschränken der Ausgabe auf die am besten/schlechtesten angepassten Modelle

Abbildung 3-4
Zeitreihenmodell zuweisen, Registerkarte "Ausgabefilter"



Die Registerkarte "Ausgabefilter" enthält Optionen, mit denen sowohl Tabellen- als auch Diagrammausgabe auf eine Untergruppe von Modellen beschränkt werden kann. Sie können die Ausgabe mithilfe von Anpassungskriterien auf die am besten und/oder am schlechtesten angepassten Modelle beschränken. In der Standardeinstellung sind alle Modelle in der Ausgabe enthalten. Sofern die Modellparameter nicht erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte "Modelle"), stammen die zum Filtern der Modelle verwendeten Werte der Anpassungsmaße aus der Modelldatei und entsprechen den Daten, die bei der Entwicklung (oder letzten Aktualisierung) des jeweiligen Modells verwendet wurden.

Am besten angepaßte Modelle. Aktivieren Sie diese Option, um die am besten angepassten Modelle in die Ausgabe einzuschließen. Wählen Sie ein Maß für Anpassungsgüte und geben Sie die Anzahl von Modellen an, die eingeschlossen werden sollen. Die Auswahl dieser Option hindert Sie nicht daran, auch die am schlechtesten angepassten Modelle auszuwählen. Die Ausgabe würde in diesem Fall sowohl die am schlechtesten als auch die am besten angepassten Modelle enthalten.

- **Feste Anzahl von Modellen.** Gibt an, dass Ergebnisse für die *n* am besten angepassten Modelle angezeigt werden. Wenn die Anzahl die Gesamtanzahl von Modellen überschreitet, werden alle Modelle angezeigt.
- **Prozentsatz der Gesamtanzahl von Modellen.** Gibt an, dass Ergebnisse für Modelle angezeigt werden, deren Werte für die Anpassungsgüte für alle Modelle in den obersten *n* Prozent liegen.

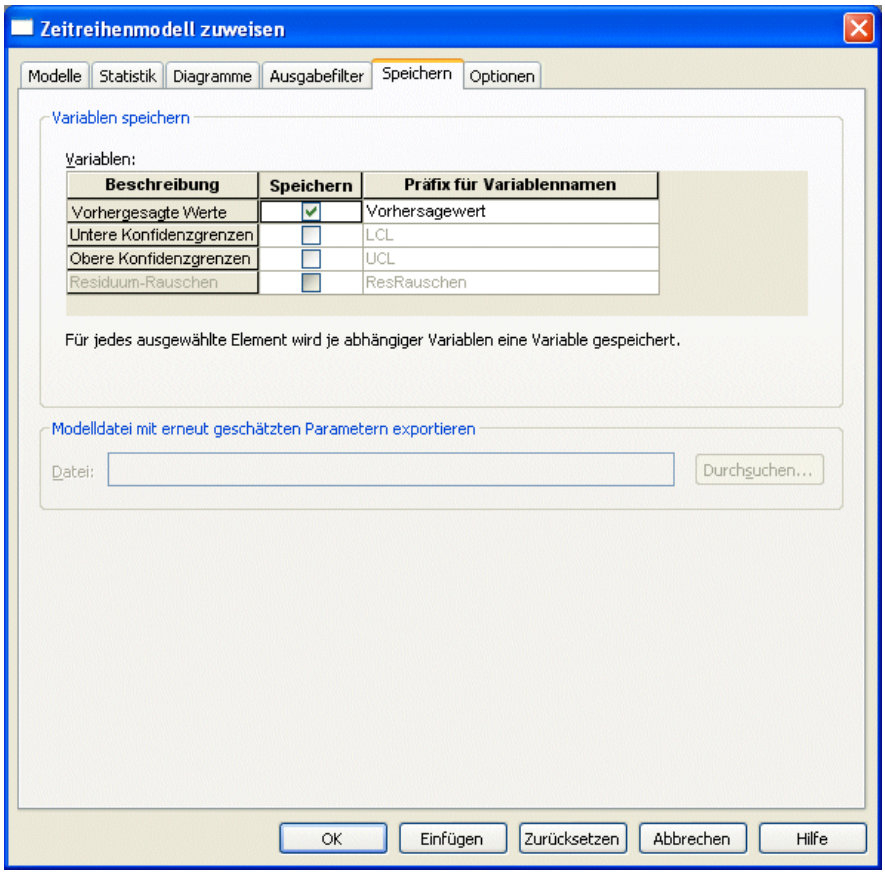
Am schlechtesten angepaßte Modelle. Aktivieren Sie diese Option, um die am schlechtesten angepassten Modelle in die Ausgabe einzuschließen. Wählen Sie ein Maß für Anpassungsgüte und geben Sie die Anzahl von Modellen an, die eingeschlossen werden sollen. Die Auswahl dieser Option hindert Sie nicht daran, auch die am besten angepassten Modelle auszuwählen. Die Ausgabe würde in diesem Fall sowohl die am besten als auch die am schlechtesten angepassten Modelle enthalten.

- **Feste Anzahl von Modellen.** Gibt an, dass Ergebnisse für die *n* am schlechtesten angepassten Modelle angezeigt werden. Wenn die Anzahl die Gesamtanzahl von Modellen überschreitet, werden alle Modelle angezeigt.
- **Prozentsatz der Gesamtanzahl von Modellen.** Gibt an, dass Ergebnisse für Modelle angezeigt werden, deren Werte für die Anpassungsgüte für alle Modelle in den untersten *n* Prozent liegen.

Maß für Anpassungsgüte. Wählen Sie das Maß für die Anpassungsgüte, anhand dessen die Modelle gefiltert werden sollen. Der Standardwert ist *R*-Quadrat.

Speichern von Modellvorhersagen und Modellspezifikationen

Abbildung 3-5
Zeitreihenmodell zuweisen, Registerkarte "Speichern"



Auf der Registerkarte "Speichern" können Sie festlegen, dass Modellvorhersagen als neue Variablen in der Arbeitsdatei gespeichert werden sollen. Außerdem können die Modellspezifikationen in einer externen Datei im XML-Format gespeichert werden.

Variablen speichern. Sie können Modellvorhersagen, Konfidenzintervalle und Residuen als neue Variablen in der Arbeitsdatei speichern. Jedes Modell erzeugt eine eigene Gruppe neuer Variablen. Wenn die Vorhersageperiode die Länge der abhängigen Zeitreihe überschreitet, die dem Modell zugewiesen ist, werden neue Fälle hinzugefügt. Sofern die Modellparameter nicht erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte "Modelle"), werden die vorhergesagten Werte und Konfidenzgrenzen nur für die Vorhersageperiode erstellt. Wenn die neuen Variablen gespeichert werden sollen, aktivieren Sie jeweils das zugehörige Kontrollkästchen "Speichern". In der Standardeinstellung werden keine neuen Variablen gespeichert.

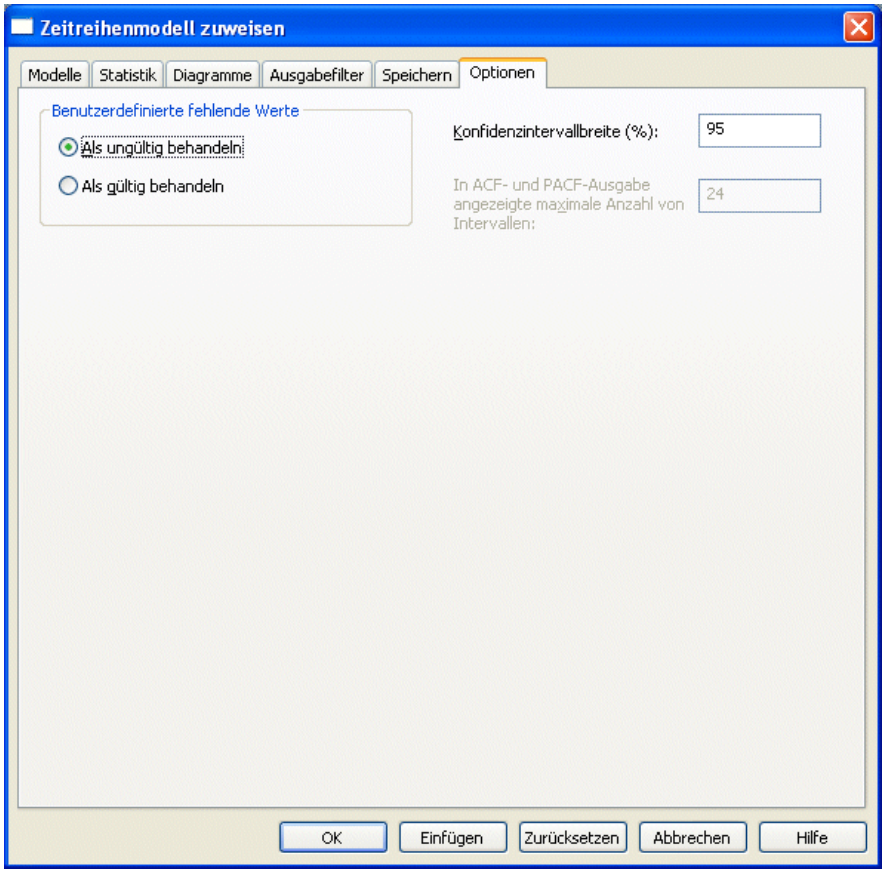
- **Vorhergesagte Werte.** Die vom Modell vorhergesagten Werte.
- **Untere Konfidenzgrenzen.** Die unteren Konfidenzgrenzen für die vorhergesagten Werte.
- **Obere Konfidenzgrenzen.** Die oberen Konfidenzgrenzen für die vorhergesagten Werte.

- **Residuum-Rauschen.** Die Modellresiduen. Bei der Transformation von abhängigen Variablen (z. B mit natürlichem Logarithmus) sind dies die Residuen für die transformierte Reihe. Diese Option ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).
- **Präfix für Variablennamen.** Geben Sie Präfixe für die Namen der neuen Variablen an oder übernehmen Sie die Standardpräfixe. Variablennamen bestehen aus dem Präfix, dem Namen der zugewiesenen abhängigen Variablen und einem Modellbezeichner. Beim Auftreten von Namenskonflikten wird der Variablenname ggf. erweitert. Das Präfix muss den Regeln für gültige Variablennamen entsprechen.

Modelldatei mit erneut geschätzten Parametern exportieren. Modellspezifikationen mit erneut geschätzten Parametern und Statistiken für Anpassungsgüte werden in die angegebene XML-Datei exportiert. Diese Option ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).

Optionen

Abbildung 3-6
Zeitreihenmodell zuweisen, Registerkarte “Optionen”



Auf der Registerkarte “Optionen” können Sie die Behandlung fehlender Werte, die Breite des Konfidenzintervalls und die für die Autokorrelationen angezeigte Anzahl von Intervallen festlegen.

Benutzerdefinierte fehlende Werte. Mit diesen Optionen wird die Behandlung benutzerdefinierter fehlender Werte gesteuert.

- **Als ungültig behandeln.** Benutzerdefiniert fehlende Werte werden wie systemdefiniert fehlende Werte behandelt.
- **Als gültig behandeln.** Benutzerdefinierte fehlende Werte werden als gültige Werte behandelt.

Richtlinie für fehlende Werte. Die folgenden Regeln gelten für die Behandlung fehlender Werte. Sie gelten auch für systemdefinierte fehlende Wert und als ungültig behandelte benutzerdefinierte fehlende Werte:

- Fälle mit fehlenden Werten einer abhängigen Variablen, die innerhalb der Schätzperiode liegen, werden in das Modell eingeschlossen. Die genaue Behandlung des fehlenden Werts hängt von der Schätzmethode ab.
- Bei ARIMA-Modellen wird eine Warnung ausgegeben, wenn ein Prädiktor innerhalb der Schätzperiode fehlende Werte aufweist. Modelle mit Prädiktoren werden nicht erneut geschätzt.
- Wenn unabhängige Variablen innerhalb der Vorhersageperiode fehlende Werte aufweisen, gibt die Prozedur eine Warnung aus und führt die Vorhersage auf der Grundlage der vorhandenen Werte aus.

Konfidenzintervallbreite (%). Konfidenzintervalle werden für die Modellvorhersagen und Residuen-Autokorrelationen berechnet. Es kann ein beliebiger positiver Wert unter 100 angegeben werden. In der Standardeinstellung wird ein Konfidenzintervall von 95 % verwendet.

In ACF- und PACF-Ausgabe angezeigte maximale Anzahl von Intervallen. Sie können die Höchstanzahl von Intervallen festlegen, die in Tabellen und Diagrammen für Autokorrelationen und partielle Autokorrelationen angezeigt werden. Diese Option ist nur verfügbar, wenn die Modellparameter erneut geschätzt werden (Erneut aus den Daten schätzen auf der Registerkarte “Modelle”).

Zusätzliche Funktionen beim Befehl TSAPPLY

Ihnen stehen zusätzliche Funktionen zur Verfügung, wenn Sie Ihre Auswahl in ein Syntax-Fenster einfügen und die resultierende Befehlssyntax für den Befehl `TSAPPLY` bearbeiten. Mit der Befehlssyntax-Sprache verfügen Sie über folgende Möglichkeiten:

- Festlegen, dass der Arbeitsdatei nur eine Untergruppe der Modelle in einer Modelldatei zugewiesen werden (mit den Schlüsselwörtern `DROP` und `KEEP` im Unterbefehl `MODEL`).
- Den Daten Modelle aus zwei oder mehr Modelldateien zuweisen (mit dem Unterbefehl `MODEL`). So könnte eine Modelldatei beispielsweise Modelle für Zeitreihen enthalten, die Verkaufsstückzahlen repräsentieren, und eine andere Modelle für Zeitreihen, die für Verkaufserlöse stehen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
4

Saisonale Zerlegung

Die Prozedur “Saisonale Zerlegung” zerlegt eine Zeitreihe in eine saisonale Komponente, eine kombinierte Trend- und Zykluskomponente sowie eine “Fehler”-Komponente. Die Prozedur ist eine Umsetzung der Census-I-Methode, die auch als “Verhältnis zum gleitenden Durchschnitt” bekannt ist.

Beispiel. Ein Wissenschaftler interessiert sich für die Analyse monatlicher Messungen des Ozongehalts an einer bestimmten Wetterstation. Ziel ist es herauszufinden, ob bei den Daten ein Trend vorliegt. Um einen etwaigen wirklichen Trend aufzudecken, muss der Wissenschaftler zunächst die Schwankungen in den Messergebnissen berücksichtigen, die auf saisonale Effekte zurückzuführen sind. Mit der Prozedur “Saisonale Zerlegung” können etwaige systematische saisonale Schwankungen entfernt werden. Die Trendanalyse wird dann auf eine saisonbereinigte Zeitreihe angewendet.

Statistiken. Die Gruppe der saisonalen Faktoren.

Daten. Die Variablen müssen numerisch sein.

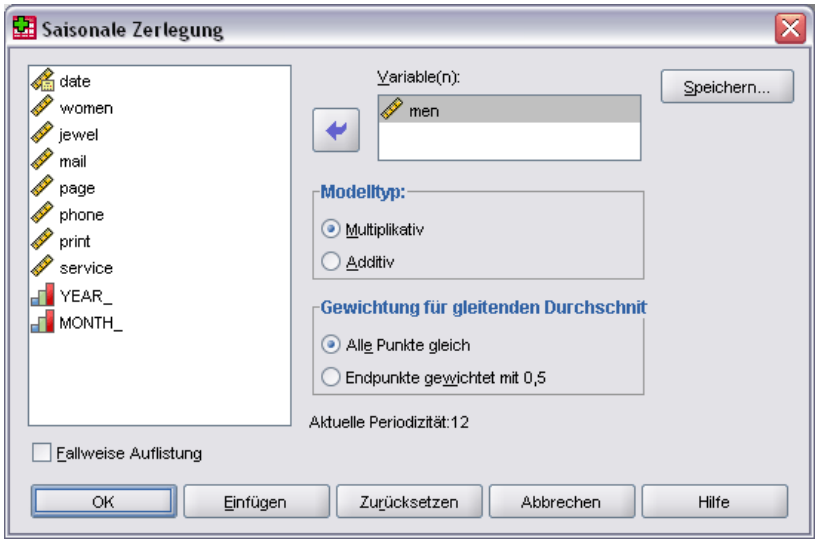
Annahmen. Die Variablen dürfen keine eingebetteten fehlenden Daten enthalten. Es muss mindestens eine periodische Datumskomponente definiert sein.

Schätzen von saisonalen Faktoren

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Saisonale Zerlegung...

Abbildung 4-1
Dialogfeld "Saisonale Zerlegung"



- Wählen Sie eine oder mehrere Variablen aus der Liste der verfügbaren Variablen und verschieben Sie sie in die Liste "Variable(n)". Achten Sie darauf, dass die Liste nur numerische Variablen enthält.

Modelltyp. Die Prozedur "Saisonale Zerlegung" bietet zwei verschiedene Ansätze für die Modellierung der saisonalen Faktoren: multiplikativ und additiv.

- **Multiplikativ.** Die saisonale Komponente ist ein Faktor, mit dem die saisonal bereinigte Zeitreihe multipliziert wird, um die ursprüngliche Zeitreihe zu erhalten. Tatsächlich werden mit dem Modul Trends saisonale Komponenten geschätzt, die proportional zum Gesamtniveau der Zeitreihe sind. Beobachtungen ohne saisonale Variation haben eine saisonale Komponente von 1.
- **Additiv.** Die saisonalen Anpassungen werden zu der saisonbereinigten Zeitreihe addiert, um die beobachteten Werte zu erhalten. Durch die Saisonbereinigung soll der saisonale Effekt aus einer Zeitreihe entfernt werden, sodass möglicherweise durch die saisonale Komponente verdeckte interessante Eigenschaften der Zeitreihe betrachtet werden können. Tatsächlich werden im Modul Trends saisonale Komponenten geschätzt, die nicht vom Niveau der Zeitreihe abhängen. Beobachtungen ohne saisonale Variation haben eine saisonale Komponente von 0.

Gewichtung für gleitenden Durchschnitt. Mit den Optionen im Gruppenfeld "Gewichtung für gleitenden Durchschnitt" können Sie festlegen, wie die Zeitreihen beim Berechnen der gleitenden Durchschnitte behandelt werden sollen. Diese Optionen sind nur verfügbar, wenn die Periodizität der Zeitreihen regelmäßig ist. Falls die Periodizität unregelmäßig ist, werden alle Punkte gleich gewichtet.

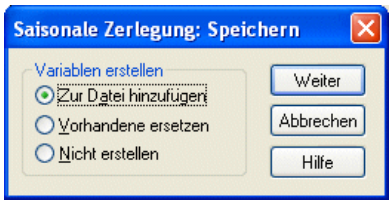
- **Alle Punkte gleich.** Gleitende Durchschnitte werden mit einer Spanne berechnet, die gleich der Periodizität ist, und so, dass alle Punkte das gleiche Gewicht erhalten. Diese Methode wird immer dann verwendet, wenn die Periodizität ungerade ist.
- **Endpunkte gewichtet mit 0,5.** Gleitende Durchschnitte für Datenreihen mit gerader Periodizität werden mit einer Spanne berechnet, die gleich der Periodizität plus 1 ist; dabei werden die Endpunkte der Spanne mit 0,5 gewichtet.

Die folgenden Optionen sind verfügbar:

- Klicken Sie auf Speichern, um anzugeben, wie neue Variablen gespeichert werden sollen.

Saisonale Zerlegung: Speichern

Abbildung 4-2
Dialogfeld "Saisonale Zerlegung: Speichern"



Variablen erstellen. Hier können Sie auswählen, wie neue Variablen behandelt werden sollen.

- **Zur Datei hinzufügen.** Die neuen Zeitreihen, die durch die saisonale Zerlegung erstellt wurden, werden als reguläre Variablen in der Arbeitsdatei gespeichert. Variablennamen werden aus einem aus drei Buchstaben bestehenden Präfix, einem Unterstrich und einer Zahl gebildet.
- **Vorhandene ersetzen.** Die neuen Zeitreihen, die durch die saisonale Zerlegung erstellt wurden, werden als temporäre Variablen in der Arbeitsdatei gespeichert. Gleichzeitig werden alle eventuell vorhandenen temporären Variablen entfernt, die durch Befehle des Moduls Trends erzeugt wurden. Variablennamen werden aus einem aus drei Buchstaben bestehenden Präfix, einer Raute (#) und einer Zahl gebildet.
- **Nicht erstellen.** Die neuen Zeitreihen werden nicht zur Arbeitsdatei hinzugefügt.

Benennung der neuen Variablen

Die Prozedur "Saisonale Zerlegung" erstellt vier neue Variablen (Zeitreihen) mit den folgenden dreibuchstabigen Präfixen für die einzelnen Reihen:

SAF. *Faktoren für die Saisonbereinigung (engl. Seasonal Adjustment Factors).* Diese Werte geben die Auswirkung der einzelnen Perioden auf das Niveau der Zeitreihe an.

SAS. *Saisonbereinigte Zeitreihe (engl. Seasonally Adjusted Series).* Dies sind die Werte, die sich nach dem Entfernen saisonaler Schwankungen einer Zeitreihe ergeben.

STC. *Geglättete Trend-Zyklus-Komponenten (engl. Smoothed Trend-Cycle Components).* Diese Werte zeigen den Trend und das zyklische Verhalten in einer Zeitreihe an.

ERR. *Residuen- oder fehlerhafte Werte (engl. Error).* Die Werte, die nach dem Entfernen der saisonalen, Trend- und Zyklus-Komponenten verbleiben.

Zusätzliche Funktionen beim Befehl SEASON

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Angabe einer etwaigen Periodizität im Befehl `SEASON`, anstatt eine der Alternativen der Prozedur “Datum definieren” auszuwählen.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Kapitel
5

Spektraldiagramme

Die Prozedur “Spektraldiagramme” wird zur Identifizierung von periodischem Verhalten bei Zeitreihen verwendet. Anstatt die Schwankungen von einem Zeitpunkt zum nächsten zu analysieren, werden die Schwankungen der Zeitreihe als Ganzes in periodischen Komponenten unterschiedlicher Frequenzen analysiert. Glatte Zeitreihen weisen stärkere periodische Komponenten bei niedrigen Frequenzen auf; durch Zufallsschwankungen (“weißes Rauschen”) wird die Komponentenstärke über alle Frequenzen verteilt.

Zeitreihen, die fehlende Daten enthalten, können mit dieser Prozedur nicht analysiert werden.

Beispiel. Die Neubauquote ist ein wichtiger Standardwert für die gesamtwirtschaftliche Lage. Die Daten für Baubeginne weisen typischerweise eine starke saisonale Komponente auf. Sind jedoch auch längere Zyklen in den Daten zu finden, deren sich die Analytiker bei der Auswertung der aktuellen Zahlen bewusst sein müssen?

Statistiken. Sinus- und Kosinus-Transformationen, Periodogramm-Wert und Spektraldichteschätzer für jede Frequenz- bzw. Periodenkomponente. Wenn eine bivariate Analyse ausgewählt ist: Real- und Imaginärteile des Kreuzperiodogramms, Kospektraldichte, Quadraturspektrum, Quadrierte Kohärenz und Phasenspektrum für jede Frequenz- bzw. Periodenkomponente.

Diagramme. Für univariate und bivariate Analysen: Periodogramm und Spektraldichte. Für bivariate Analysen: quadrierte Kohärenz, Quadraturspektrum, Kreuzamplitude, Kospektraldichte, Phasenspektrum und Gewinn.

Daten. Die Variablen müssen numerisch sein.

Annahmen. Die Variablen dürfen keine eingebetteten fehlenden Daten enthalten. Die zu analysierende Zeitreihe muss stationär sein und ein Mittelwert ungleich 0 muss aus der Zeitreihe subtrahiert werden.

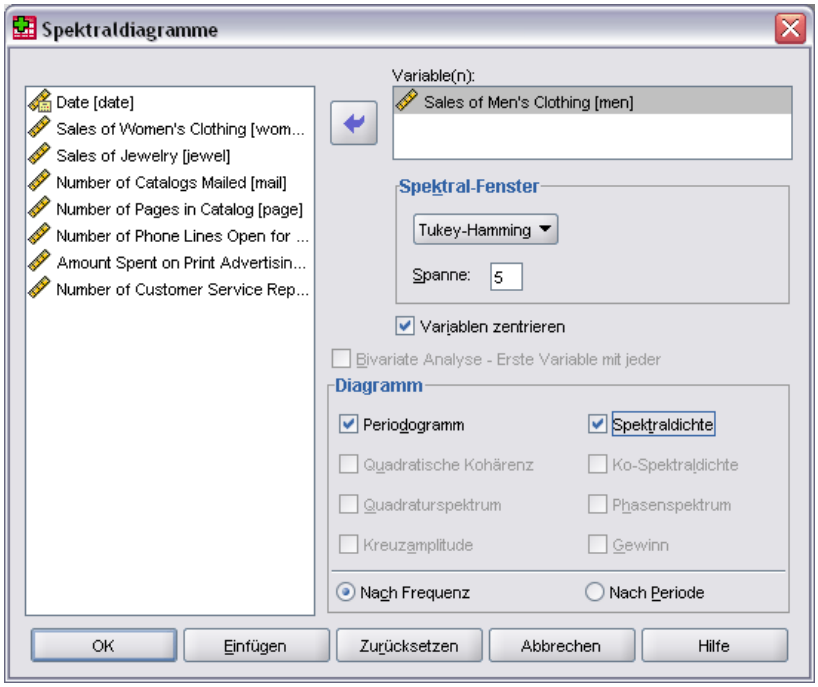
- **Stationär.** Eine Bedingung, die von Zeitreihen eingehalten werden muss, an die ARIMA-Modelle angepasst werden sollen. Reine MA-Reihen sind stationär, aber AR- und ARMA-Reihen können auch nichtstationär sein. Eine stationäre Zeitreihe besitzt einen konstanten Mittelwert und eine konstante Varianz über die Zeit.

Berechnen einer Spektralanalyse

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analyse
  Zeitreihen
    Spektralanalyse...

Abbildung 5-1
Dialogfeld "Spektraldiagramme"



- Wählen Sie eine oder mehrere Variablen aus der Liste der verfügbaren Variablen und verschieben Sie sie in die Liste "Variable(n)". Achten Sie darauf, dass die Liste nur numerische Variablen enthält.

- Wählen Sie eine der Optionen im Gruppenfeld "Spektral-Fenster", um festzulegen, wie das Periodogramm geglättet werden soll, um eine Spektraldichteschätzung zu erhalten. Folgende Glättungsoptionen stehen zur Verfügung: "Tukey-Hamming", "Tukey", "Parzen", "Bartlett", "Daniell (Einheit)" und "Keine".
  - **Tukey-Hamming.** Die Gewichte lauten: Wk = 0,54Dp(2 pi fk) + 0,23Dp (2 pi fk + pi/p) + 0,23Dp (2 pi fk - pi/p), für k = 0, ..., p. Dabei ist p der ganzzahlige Anteil der Hälfte der Spannweite und Dp ist der Dirichlet-Kern der Ordnung p.
  - **Tukey.** Die Gewichte lauten: Wk = 0,5Dp(2 pi fk) + 0,25Dp (2 pi fk + pi/p) + 0,25Dp(2 pi fk - pi/p), für k = 0, ..., p. Hierbei ist p der ganzzahlige Teil der Spannweitenhälfte und Dp ist der Dirichlet-Kern der Ordnung p.
  - **Parzen.** Die Gewichtungen sind Wk = 1/p(2 + cos(2 pi fk)) (F[p/2] (2 pi fk))**2, für k= 0, ... p. Dabei ist p der ganzzahlige Teil der halben Spannweite und F[p/2] ist der Fejérsche Kern der Ordnung p/2.
  - **Bartlett.** Die Form eines Spektralfensters, bei dem die Gewichte der oberen Hälfte gleich Wk = Fp (2*pi*fk), für k = 0, ... p sind. Dabei bezeichnet p den ganzzahligen Anteil der halben Spannweite. Fp ist der Fejér'sche Kern der Ordnung p. Die untere Hälfte des Fensters ist symmetrisch zur oberen.

- **Daniell (Einheit).** Die Form eines Spektralfensters, für das alle Gewichtungen gleich 1 sind.
- **Keine.** Keine Glättung. Wenn diese Option ausgewählt wird, entspricht die Schätzung der Spektraldichte dem Periodogramm.

Spanne. Der Bereich aufeinanderfolgender Werte, für den die Glättung durchgeführt wird. In der Regel werden ungerade ganze Zahlen verwendet. Bei größeren Spannweiten ist die Glättung der Spektraldichte-Diagramme größer als bei kleineren Spannweiten.

Variablen zentrieren. Passt die Zeitreihe so an, dass sie vor der Berechnung des Spektrums den Mittelwert 0 hat, sodass der eventuell mit dem Mittelwert zusammenhängende große Term entfernt wird.

Bivariate Analyse — Erste Variable mit jeder. Wenn Sie zwei oder mehr Variablen ausgewählt haben, können Sie diese Option aktivieren, um eine bivariate Spektralanalyse berechnen zu lassen.

- Die erste Variable in der Liste “Variable(n)” wird als unabhängige Variable behandelt und alle übrigen Variablen gelten als abhängige Variablen.
- Jede auf die erste Zeitreihe folgende Zeitreihe wird unabhängig von anderen festgelegten Zeitreihen mit der ersten Zeitreihe analysiert. Außerdem werden univariate Analysen jeder Zeitreihe durchgeführt.

Diagramm. Periodogramm und Spektraldichte sind sowohl für univariate und bivariate Analysen verfügbar. Alle anderen Optionen stehen nur für bivariate Analysen zur Verfügung.

- **Periodogramm.** Ein ungeglättetes Diagramm der Spektralamplitude (dargestellt auf einer logarithmischen Skala) in Abhängigkeit von Frequenz oder Periode. Eine niederfrequente Streuung charakterisiert eine glatte Zeitreihe. Eine gleichmäßig über alle Frequenzen verteilte Streuung ist ein Zeichen für "weißes Rauschen".
- **Quadratische Kohärenz.** Das Produkt der Zunahmewerte der beiden Zeitreihen.
- **Quadraturspektrum.** Der Imaginärteil des Kreuzperiodogramms; er ist ein Maß für die Korrelation der phasenverschobenen Frequenzkomponenten der beiden Zeitreihen. Die Komponenten sind um pi/2 phasenverschoben.
- **Kreuzamplitude.** Die Wurzel aus der Summe von quadrierter Kospektraldichte und quadriertem Quadraturspektrum.
- **Spektraldichte.** Ein Periodogramm, das geglättet wurde, um unregelmäßige Variationen zu entfernen.
- **Kospektraldichte.** Der Realteil des Kreuzperiodogramms; er ist ein Maß für die Korrelation der phasengleichen Frequenzkomponenten der beiden Zeitreihen.
- **Phasenspektrum.** Ein Maß dafür, inwieweit jede Frequenzkomponente einer Zeitreihe der anderen Zeitreihe voraus- oder nachläuft.
- **Zunahme (Gewinn).** Der Quotient aus Kreuzamplitude und Spektraldichte für eine der Zeitreihen. Jede der beiden Zeitreihen hat einen eigenen Zunahmewert.

Nach Frequenz. Alle Diagramme werden nach der Frequenz erstellt. Der Frequenzbereich reicht von 0 (der konstante Term oder Mittelwertterm) bis 0,5 (der Term für einen Zyklus von zwei Beobachtungen).

Nach Periode. Alle Diagramme werden nach der Periode erstellt. Der Periodenbereich reicht von 2 (der Term für einen Zyklus von zwei Beobachtungen) bis zur Anzahl der Beobachtungen (der konstante Term oder Mittelwertterm). Die Periode wird auf einer logarithmischen Skala dargestellt.

Zusätzliche Funktionen beim Befehl SPECTRA

Mit der Befehlssyntax-Sprache verfügen Sie außerdem über folgende Möglichkeiten:

- Die berechneten Spektralanalysevariablen können für eine spätere Verwendung in der Arbeitsdatei gespeichert werden.
- Für das Spektral-Fenster können benutzerdefinierte Gewichtungen festgelegt werden.
- Es können Diagramme sowohl nach Frequenz als auch nach Periode erstellt werden.
- Es kann eine vollständige Auflistung jedes im Diagramm angezeigten Werts ausgegeben werden.

Vollständige Informationen zur Syntax finden Sie in der *Command Syntax Reference*.

Teil II: Beispiele

Kapitel 6

Prognosen für große Datenmengen mit dem Expert Modeler

Ein Analyst eines Breitband-Providers soll eine Prognose über die Vertragsabschlüsse mit Kunden erstellen, um die Nutzung der Bandbreite vorhersagen zu können. Es werden Prognosen für alle 85 regionalen Märkte benötigt, die zusammen den gesamten Kundenstamm ergeben. Die früheren monatlichen Daten befinden sich in *broadband_1.sav*. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108.

In diesem Beispiel werden Sie den Expert Modeler verwenden, um für die kommenden drei Monaten für jeden der 85 regionalen Märkte Prognosen anzufertigen, wobei die erzeugten Modelle in einer externe XML-Datei gespeichert werden. Wenn Sie fertig sind, empfiehlt es sich möglicherweise, das nächste Beispiel, Neuprognosen für große Datenmengen durch das Zuweisen gespeicherter Modelle in Kapitel 7 auf S. 60, durchzuarbeiten. In diesem Beispiel werden die gespeicherten Modelle auf ein aktualisiertes Daten-Set angewendet, um die Prognosen auf drei weitere Monate auszuweiten, ohne das Modell erneut zu erstellen.

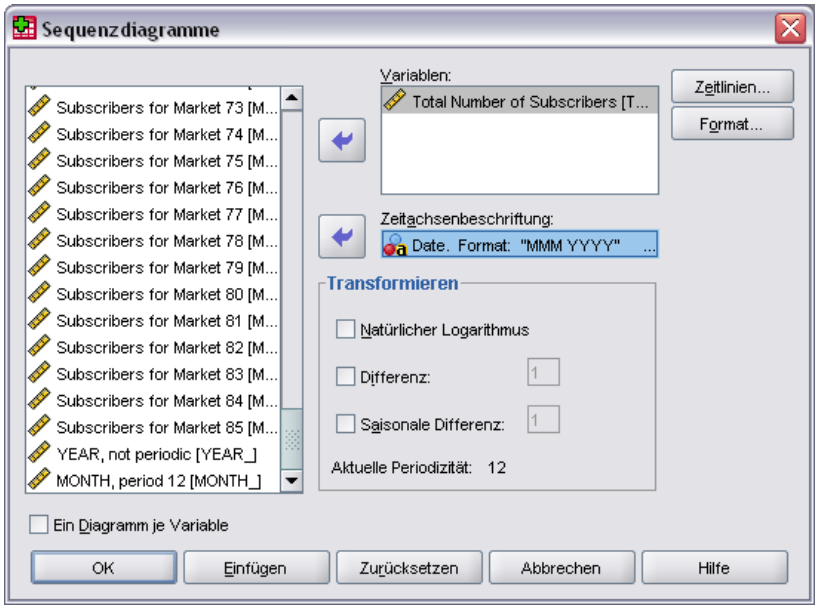
Untersuchen der Daten

Es empfiehlt sich grundsätzlich, ein Gefühl für die Natur der Daten zu entwickeln, bevor Sie das Modell erstellen. Liegen in den Daten saisonale Variationen vor? Auch wenn der Expert Modeler für jede Zeitreihe das beste saisonale oder nichtsaisonale Modell findet, können Sie Ergebnisse häufig schneller berechnen, indem Sie die Suche auf nichtsaisonale Modelle begrenzen, sofern in den Daten keine Saisonalität vorliegt. Ohne die Daten für jeden der 85 regionalen Märkte zu untersuchen, lässt sich ein grobes Bild der Situation gewinnen, in dem die Gesamtzahl der Vertragskunden für alle Märkte grafisch dargestellt wird.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

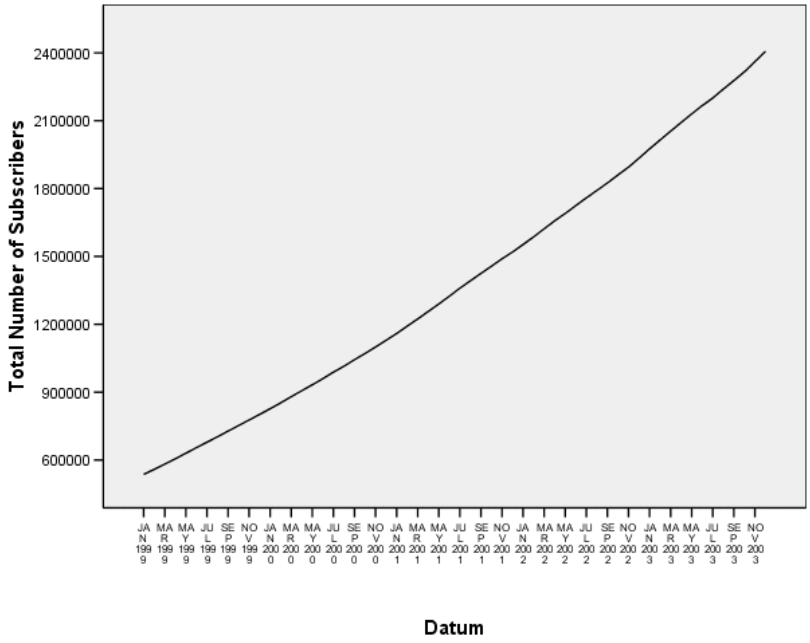
Analysieren
  Zeitreihen
    Sequenzdiagramme...

Abbildung 6-1
Dialogfeld "Sequenzdiagramme"



- Wählen Sie die Variable *Total Number of Subscribers* aus und verschieben Sie sie in die Liste "Variablen".
- Wählen Sie die Variable *Date* aus und verschieben Sie sie in das Feld "Zeitachsenbeschriftung".
- Klicken Sie auf OK.

Abbildung 6-2
Gesamtzahl der Vertragskunden für Breitbandnutzung auf allen Märkten



Die Zeitreihe zeigt einen sehr gleichmäßigen Aufwärtstrend ohne Anzeichen für saisonale Variationen. Möglicherweise weisen einzelne Zeitreihen Saisonalität auf, jedoch scheint die Saisonalität im Allgemeinen kein ausgeprägtes Merkmal der Daten zu sein. Selbstverständlich müssen Sie jede der Zeitreihen untersuchen, bevor Sie saisonale Modelle ausschließen. Sie können dann die Zeitreihen aussondern, die Saisonalität aufweisen, und diese separat modellieren. Im vorliegenden Fall zeigt die Untersuchung der 85 Zeitreihen, dass kein Markt Saisonalität aufweist.

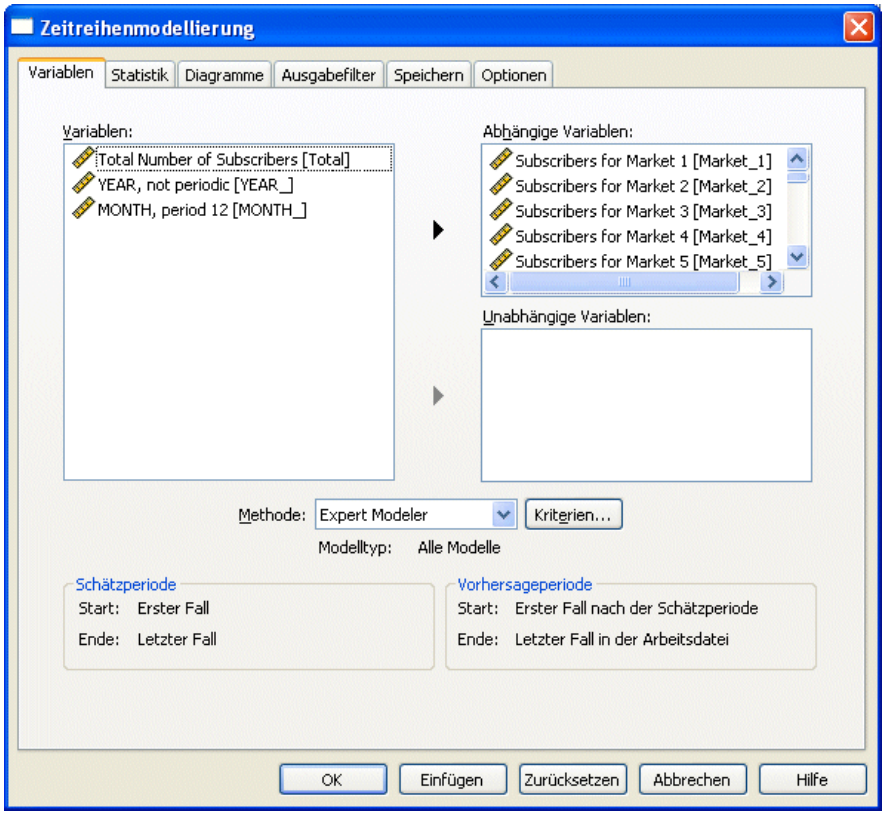
Durchführung der Analyse

So verwenden Sie den Expert Modeler:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Modelle erstellen...

Abbildung 6-3
Dialogfeld "Zeitreihenmodellierung"



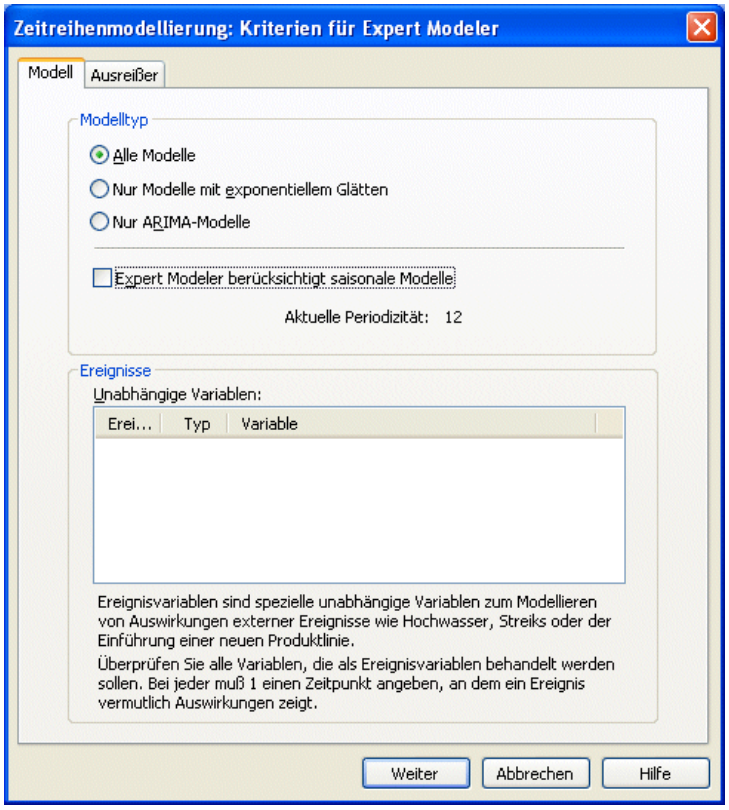
- Wählen Sie die Variablen *Subscribers for Market 1* bis *Subscribers for Market 85* als abhängige Variablen aus.
- Stellen Sie sicher, dass in der Dropdown-Liste "Methode" die Option Expert Modeler ausgewählt ist. Der Expert Modeler ermittelt automatisch das jeweils am besten angepasste Modell für jede der abhängigen Zeitreihen.

Die Gruppe der Fälle, die für die Schätzung des Modells verwendet wird, wird als **Schätzperiode** bezeichnet. In der Standardeinstellung umfasst sie sämtliche Fälle in der Arbeitsdatei. Sie können die Schätzperiode festlegen, indem Sie im Dialogfeld "Fälle auswählen" die Option Nach Zeit- oder Fallbereich auswählen. Dieses Beispiel bezieht sich auf die Standardeinstellung.

Beachten Sie auch, dass die Standard-Vorhersageperiode nach dem Ende der Schätzperiode beginnt und sich bis zum letzten Fall in der Arbeitsdatei erstreckt. Wenn die Vorhersage über den letzten Fall hinausgehen soll, müssen Sie die Vorhersageperiode ausweiten. Sie können diese Einstellung auf der Registerkarte "Optionen" vornehmen, wie später in diesem Beispiel erläutert wird.

▶ Klicken Sie auf Kriterien.

Abbildung 6-4
Dialogfeld "Kriterien für Expert Modeler", Registerkarte "Modell"


Zeitreihenmodellierung: Kriterien für Expert Modeler
Modell
Ausreißer
Modelltyp
Alle Modelle
Nur Modelle mit exponentiellem Glätten
Nur ARIMA-Modelle
Expert Modeler berücksichtigt saisonale Modelle
Aktuelle Periodizität: 12
Ereignisse
Unabhängige Variablen:
Erei...
Typ
Variable
Ereignisvariablen sind spezielle unabhängige Variablen zum Modellieren von Auswirkungen externer Ereignisse wie Hochwasser, Streiks oder der Einführung einer neuen Produktlinie.
Überprüfen Sie alle Variablen, die als Ereignisvariablen behandelt werden sollen. Bei jeder muß 1 einen Zeitpunkt angeben, an dem ein Ereignis vermutlich Auswirkungen zeigt.
Weiter
Abbrechen
Hilfe

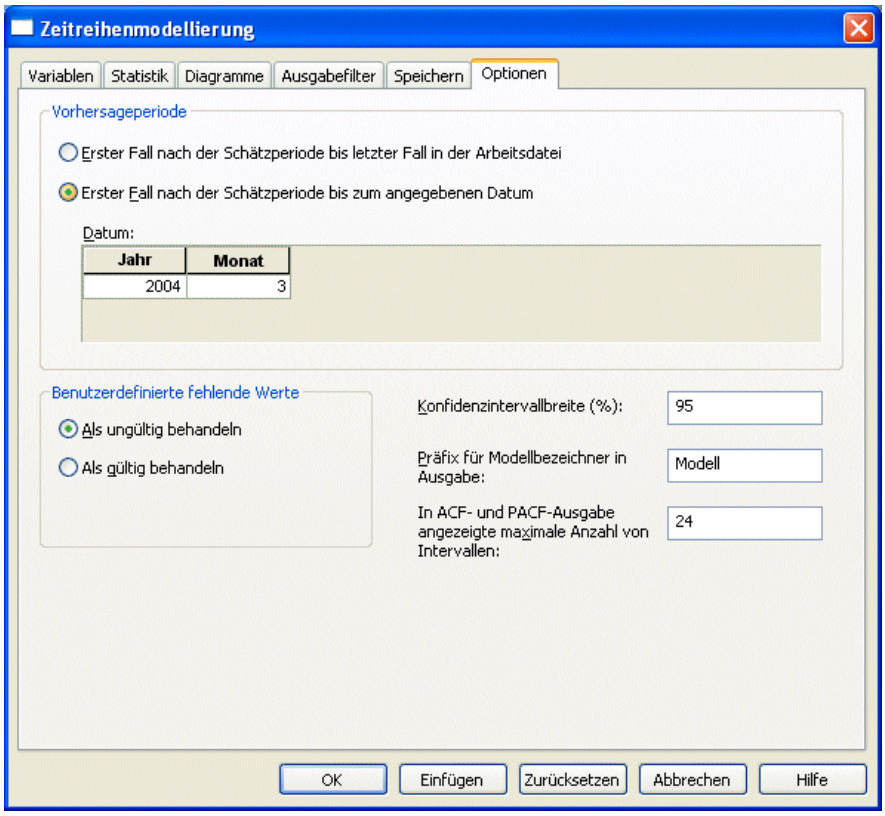
▶ Deaktivieren Sie im Gruppenfeld "Modelltyp" die Option Expert Modeler berücksichtigt saisonale Modelle.

Wenngleich die Daten monatlich erhoben werden und die aktuelle Periodizität 12 beträgt, konnte beobachtet werden, dass die Daten keine Saisonalität aufweisen. Damit besteht keine Notwendigkeit, saisonale Modelle zu berücksichtigen. Dies verringert den Raum der Modelle, der von Expert Modeler durchsucht wird, und die Zeit für die Berechnung wird deutlich verringert.

▶ Klicken Sie auf Weiter.

▶ Klicken Sie im Dialogfeld "Zeitreihenmodellierung" auf die Registerkarte Optionen.

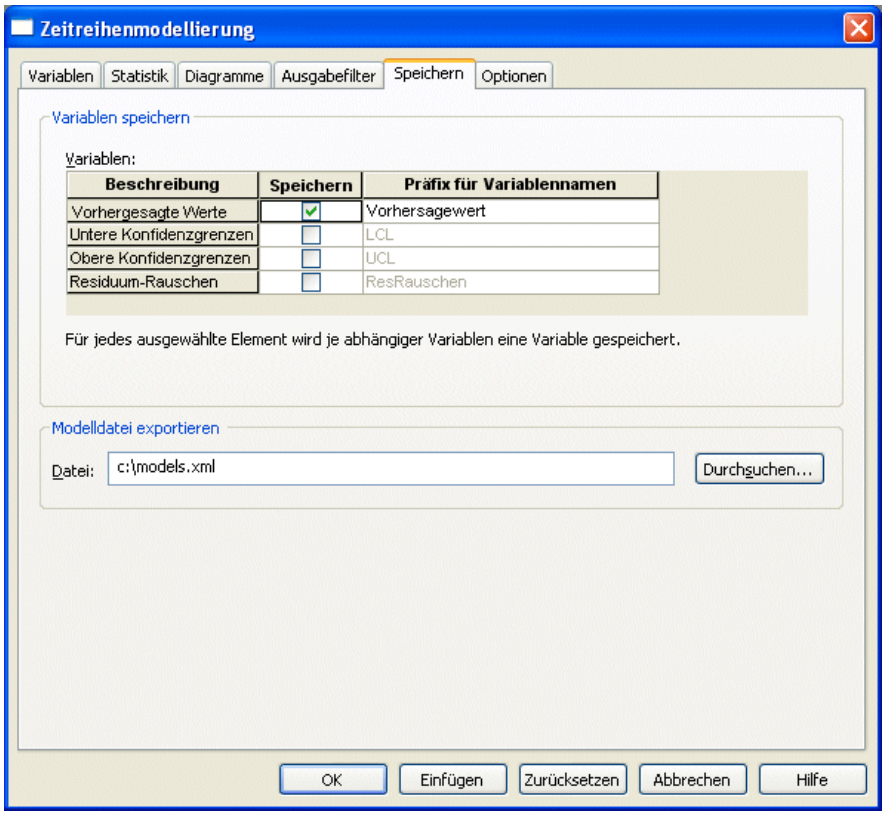
Abbildung 6-5
Zeitreihenmodellierung, Registerkarte "Optionen"



- Wählen Sie im Gruppenfeld "Vorhersageperiode" die Option Erster Fall nach der Schätzperiode bis zum angegebenen Datum aus.
- Geben Sie im Datumsgitter für das Jahr 2004 und für den Monat 3 ein.

  Das Daten-Set enthält Daten von Januar 1999 bis Dezember 2003. Mit den aktuellen Einstellungen erstreckt sich die Vorhersageperiode von Januar 2004 bis März 2004.
- Klicken Sie auf die Registerkarte Speichern.

Abbildung 6-6
Zeitreihenmodellierung, Registerkarte "Speichern"



- Aktivieren Sie in der Spalte *Speichern* den Eintrag "Vorhergesagte Werte", und behalten Sie den Standardwert *Vorhersagewert* als "Präfix für Variablennamen" bei.

  Die Modellvorhersagen werden als neue Variablen in der Arbeitsdatei gespeichert, wobei *Vorhersagewert* als Präfix für die Variablennamen verwendet wird. Sie können die Spezifikationen für jedes Modell auch in einer externen XML-Datei speichern. Dies ermöglicht Ihnen, die Modelle erneut zu verwenden, wenn Sie die Vorhersage bei der Verfügbarkeit neuer Daten ausweiten möchten.

- Klicken Sie auf der Registerkarte "Speichern" auf die Schaltfläche Durchsuchen.

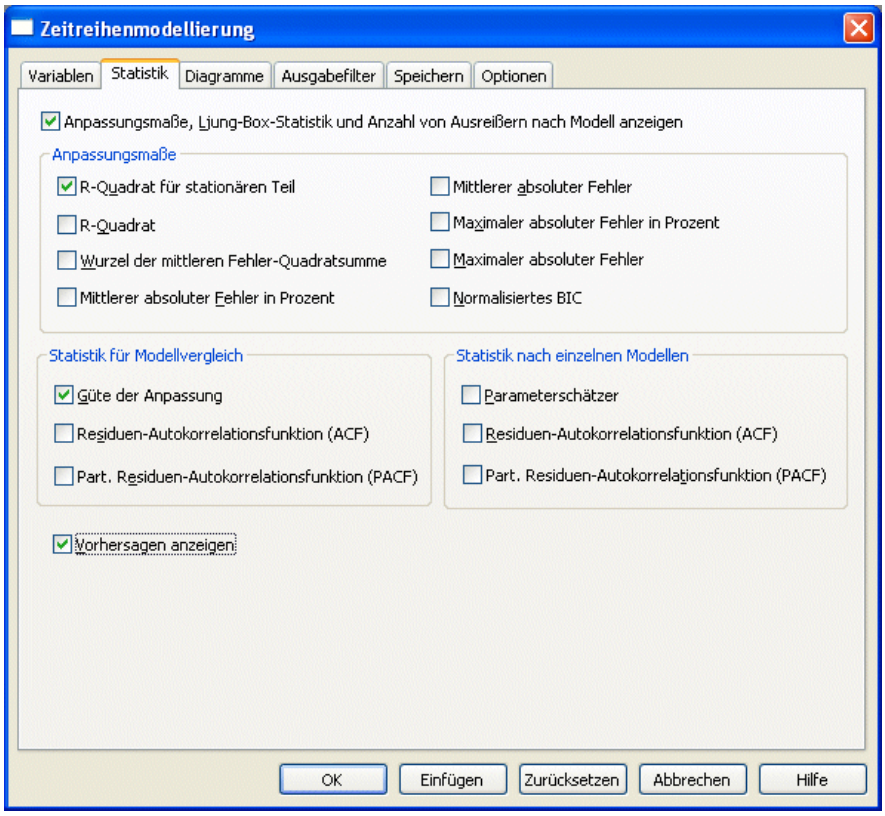
  Dadurch wird ein Standarddialogfeld zum Speichern von Dateien geöffnet.

- Wechseln Sie in den Ordner, in dem Sie die XML-Modelldatei speichern möchten. Geben Sie einen Dateinamen ein, und klicken Sie auf Speichern.

  Der Pfad für die XML-Modelldatei wird daraufhin auf der Registerkarte "Speichern" angezeigt.

- Klicken Sie auf die Registerkarte Statistik.

Abbildung 6-7
Zeitreihenmodellierung, Registerkarte "Statistik"



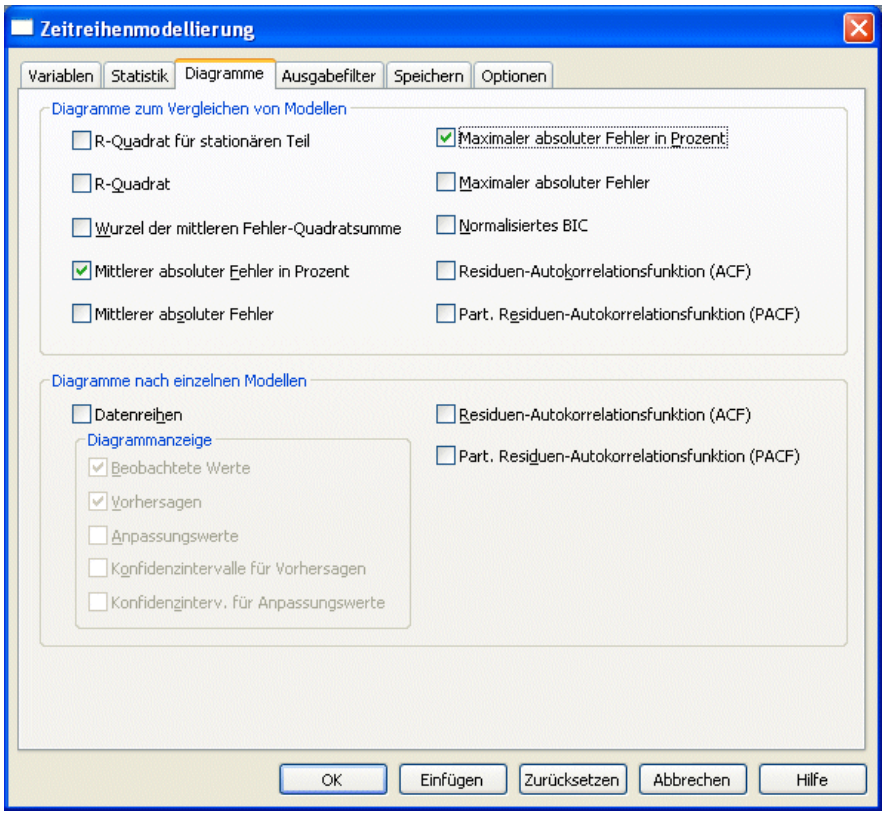
▶ Wählen Sie Vorhersagen anzeigen aus.

Mithilfe dieser Option wird für jede abhängige Variablenreihe eine Tabelle der vorhergesagten Werte erstellt. Dies ist neben der Möglichkeit, die Vorhersagen als neue Variablen zu speichern, eine weitere Option, mit der diese Werte berechnet werden können.

Bei der Standardauswahl Güte der Anpassung (im Gruppenfeld "Statistik für Modellvergleich") wird eine Tabelle mit modellübergreifend berechneten Anpassungsstatistiken erstellt, z. B. *R*-Quadrat, mittlerer absoluter Fehler in Prozent und normalisiertes BIC. Diese Tabelle bietet eine präzise Zusammenfassung dafür, wie gut das Modell an die Daten angepasst ist.

▶ Klicken Sie auf die Registerkarte Diagramme.

Abbildung 6-8
Zeitreihenmodellierung, Registerkarte "Diagramme"



▶ Deaktivieren Sie im Gruppenfeld "Diagramme nach einzelnen Modellen" die Option Datenreihen.

Hierbei werden die Zeitreihen-Diagramme für einzelne Modelle unterdrückt. Das vorliegende Beispiel soll in erster Linie erläutern, wie die Vorhersagen in neuen Variablen gespeichert werden. Das Erzeugen von Diagrammen der Vorhersagen ist nur von sekundärem Interesse.

Mit den Optionen im Gruppenfeld "Diagramme zum Vergleichen von Modellen" können verschiedene Diagramme (in Form von Histogrammen) von modellübergreifend berechneten Anpassungsstatistiken erstellt werden.

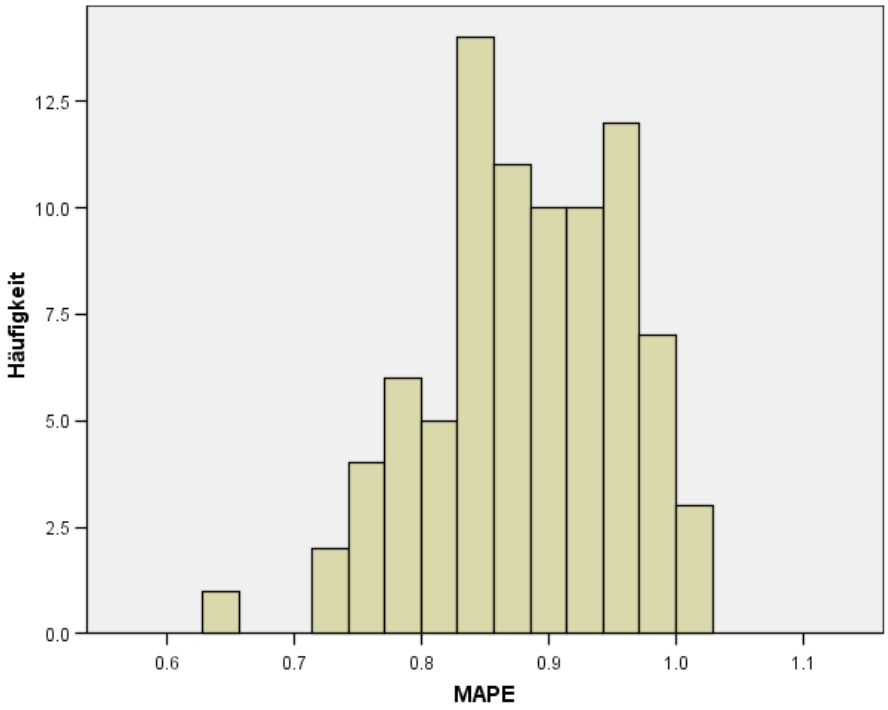
▶ Wählen Sie im Gruppenfeld "Diagramme zum Vergleichen von Modellen" die Optionen Mittlerer absoluter Fehler in Prozent und Maximaler absoluter Fehler in Prozent aus.

Der absolute Fehler in Prozent ist ein Maß dafür, wie sehr eine abhängige Zeitreihe vom Niveau abweicht, das vom Modell vorhergesagt wurde. Wenn Sie den mittleren und maximalen Prozentsatz modellübergreifend untersuchen, erhalten Sie einen Hinweis auf die Unsicherheit in Ihren Vorhersagen. Es empfiehlt sich außerdem, Auswertungsdiagramme der prozentualen statt der absoluten Fehler heranzuziehen, da die abhängigen Zeitreihen die Anzahl der Vertragskunden für Märkte mit unterschiedlichen Größen darstellen.

▶ Klicken Sie im Dialogfeld "Zeitreihenmodellierung" auf OK.

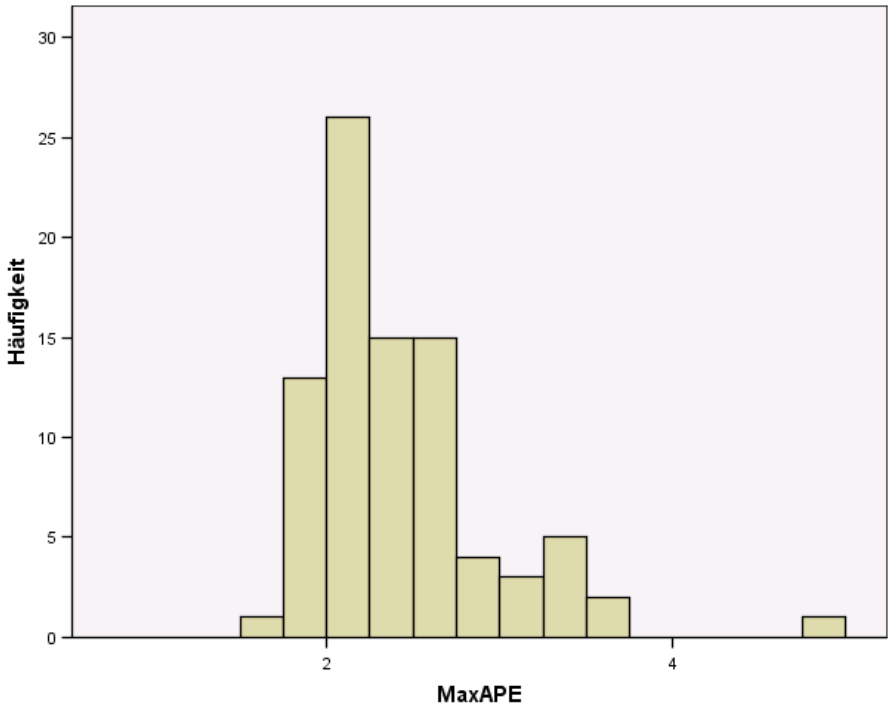
Diagramme für die Modellzusammenfassung

Abbildung 6-9
Histogramm des mittleren absoluten Fehlers in Prozent



Dieses Histogramm zeigt den modellübergreifenden mittleren absoluten Fehler in Prozent (MAPE) an. Es zeigt, dass alle Modelle eine mittlere Unsicherheit von ungefähr 1% aufweisen.

Abbildung 6-10
Histogramm des maximalen absoluten Fehlers in Prozent



Dieses Histogramm zeigt den modellübergreifenden maximalen absoluten Fehler in Prozent (MaxAPE) an. Es eignet sich dazu, den schlechtesten Fall für Ihre Vorhersage zu skizzieren. Es zeigt, dass für jedes Modell die größte Fehlerprozentzahl im Bereich von 1 bis 5% liegt. Stellen diese Werte eine akzeptable Unsicherheit dar? Dies ist eine Situation, in der Ihr Geschäftssinn ins Spiel kommt, da ein akzeptables Risiko von Fall zu Fall variiert.

Modellvorhersagen

Abbildung 6-11
Neue Variablen mit Modellvorhersagen

YEAR_	MONTH_	DATE_	Vorhersagewert_Market_1_Modell_1	Vorhersagewert_Market_2_Modell_2
2003	10	OCT 2003	11820	51084
2003	11	NOV 2003	11857	51273
2003	12	DEC 2003	11687	53082
2004	1	JAN 2004	11503	54893
2004	2	FEB 2004	11447	55856
2004	3	MAR 2004	11390	56704

Der Daten-Editor zeigt die neuen Variablen an, die die Modellvorhersagen enthalten. Auch wenn hier nur zwei gezeigt werden, sind 85 neue Variablen vorhanden: eine für jede der 85 abhängigen Zeitreihen. Die Variablennamen bestehen aus dem Standardpräfix *Vorhersagewert*, gefolgt vom Namen der verknüpften abhängigen Variable (zum Beispiel *Market_1* und einem Modellbezeichner (zum Beispiel *Modell_1*).

Dem Daten-Set wurden drei neue Fälle mit den Vorhersagen von Januar 2004 bis März 2005 und automatisch erzeugten Datenlabels hinzugefügt. Jede der neue Variablen enthält Modellvorhersagen für die Schätzperiode (Januar 1999 bis Dezember 2003). Damit bekommen Sie einen Einblick, wie gut das Modell an die bekannten Werte angepasst ist.

Abbildung 6-12
Vorhersagetabelle

Modell		JAN 2004	FEB 2004	Mär 2004
Subscribers for Market 1-Model_1	Vorhersagen	11503	11447	11390
	UCL	11686	11767	11870
	LCL	11321	11126	10910
Subscribers for Market 2-Model_2	Vorhersagen	54893	55856	56704
	UCL	55632	57195	58575
	LCL	54154	54518	54832
Subscribers for Market 3-Model_3	Vorhersagen	59656	59305	58954
	UCL	60457	60753	61158
	LCL	58856	57857	56750
Subscribers for Market 4-Model_4	Vorhersagen	18235	18424	18628
	UCL	18413	18731	19121
	LCL	18058	18116	18136

Sie können auch eine Tabelle mit den vorhergesagten Werten erstellen. Die Tabelle besteht aus den vorhergesagten Werten in der Schätzperiode, enthält jedoch im Gegensatz zu den neuen Variablen mit den Modellvorhersagen keine vorhergesagten Werte in der Schätzperiode. Die Ergebnisse sind nach Modellen geordnet und mit dem Modellnamen beschriftet. Dieser besteht aus dem Namen (oder Label) der verknüpften abhängigen Variablen gefolgt von einem Modellbezeichner, genauso wie die Namen der neue Variablen, die die Modellvorhersagen enthalten. Die Tabelle enthält auch die oberen Konfidenzgrenzen (UCL) und die unteren Konfidenzgrenzen (LCL) für die vorhergesagten Werte (95 % in der Standardeinstellung).

Sie können zwischen zwei Ansätzen für die Berechnung der vorhergesagten Werte auswählen: Speichern der Vorhersagen als neue Variablen in der Arbeitsdatei oder Erstellen einer Vorhersagetabelle. Bei beiden Ansätzen verfügen Sie über eine Vielzahl von Optionen zum Exportieren der Vorhersagen (z. B. nach Excel).

Zusammenfassung

Sie haben die Verwendung von Expert Modeler kennen gelernt, um damit Vorhersagen für mehrere Zeitreihen zu erstellen, und Sie haben die resultierenden Modelle in einer externe XML-Datei gespeichert. Im nächsten Beispiel werden Sie erfahren, wie Sie mithilfe der Prozedur “Zeitreihenmodell zuweisen” Vorhersagen bei der Verfügbarkeit neuer Daten ausweiten können, ohne dass Sie die Modelle erneut erstellen müssen.

Kapitel 7

Neuprognosen für große Datenmengen durch das Zuweisen gespeicherter Modelle

Sie haben mit der Zeitreihenmodellierung Modelle für Zeitreihendaten erstellt und erste Vorhersagen auf der Grundlage verfügbarer Daten erzeugt. Sie möchten diese Modelle für eine Erweiterung der Vorhersagen verwenden, wenn aktuellere Daten verfügbar werden, daher haben Sie die Modelle in einer externen Datei gespeichert. Nun können Sie die gespeicherten Modelle zuweisen.

Dieses Beispiel ist eine Erweiterung des vorhergehenden Beispiels, Prognosen für große Datenmengen mit dem Expert Modeler in Kapitel 6 auf S. 49, kann jedoch auch unabhängig verwendet werden. In diesem Szenario sind Sie Analytiker für einen nationalen Breitband-Provider und müssen monatliche Vorhersagen der Vertragskunden in 85 regionalen Märkten aufstellen. Sie haben bereits Modelle mit dem Expert Modeler erstellt und eine Vorhersage über drei Monate angefertigt. Das Data Warehouse wurde mit tatsächlichen Daten für die ursprüngliche Vorhersageperiode aktualisiert. Daher möchten Sie den Vorhersagehorizont mithilfe dieser Daten um weitere drei Monate erweitern.

Die aktualisierten früheren monatlichen Daten befinden sich in *broadband_2.sav* und die gespeicherten Modelle in *broadband_models.xml*. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108. Wenn Sie das vorhergehende Beispiel durchgearbeitet und eine eigene Modelldatei gespeichert haben, können Sie selbstverständlich anstelle von *broadband_models.xml* Ihre eigene Datei verwenden.

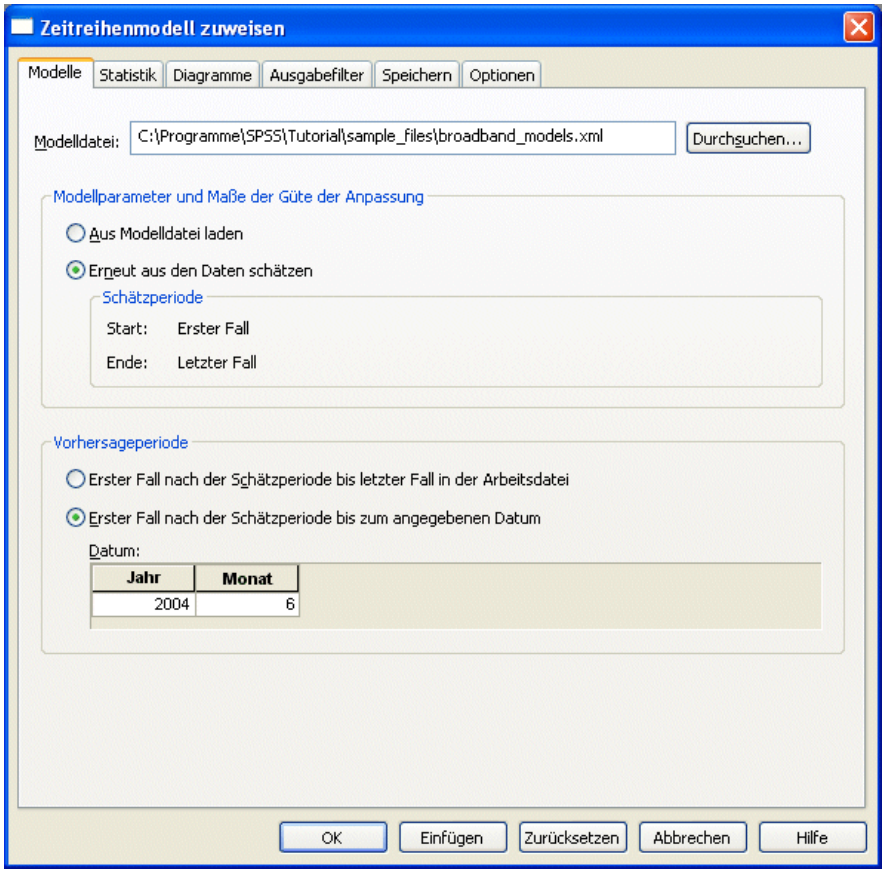
Durchführung der Analyse

So weisen Sie Modelle zu:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Modelle zuweisen...

Abbildung 7-1
Dialogfeld "Zeitreihenmodell zuweisen"



- Klicken Sie auf Durchsuchen und wechseln Sie dann zu der Datei *broadband_models.xml* und wählen Sie sie aus (alternativ können Sie auch Ihre eigene Modelldatei aus dem vorhergehenden Beispiel auswählen). Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108.

  Der Pfad von *broadband_models.xml* (oder der eigenen Modelldatei) wird nun auf der Registerkarte "Modelle" angezeigt.

- Wählen Sie Erneut aus den Daten schätzen aus.

  Um neue Werte der Zeitreihe in Vorhersagen aufzunehmen, müssen die Modellparameter mit der Prozedur "Zeitreihenmodell zuweisen" erneut geschätzt werden. Die Struktur der Modelle bleibt jedoch unverändert, sodass für die Neuschätzung wesentlich weniger Rechenzeit als beim ersten Erstellen der Modelle erforderlich ist.

  Die für die Neuschätzung verwendete Gruppe von Fällen muss die neuen Daten enthalten. Wenn Sie die Standardschätzperiode von "Erster Fall" bis "Letzter Fall" verwenden, ist dies sichergestellt. Wenn Sie als Schätzperiode einen anderen als den Standardwert festlegen müssen, können Sie im Dialogfeld "Fälle auswählen" die Option Nach Zeit- oder Fallbereich auswählen.

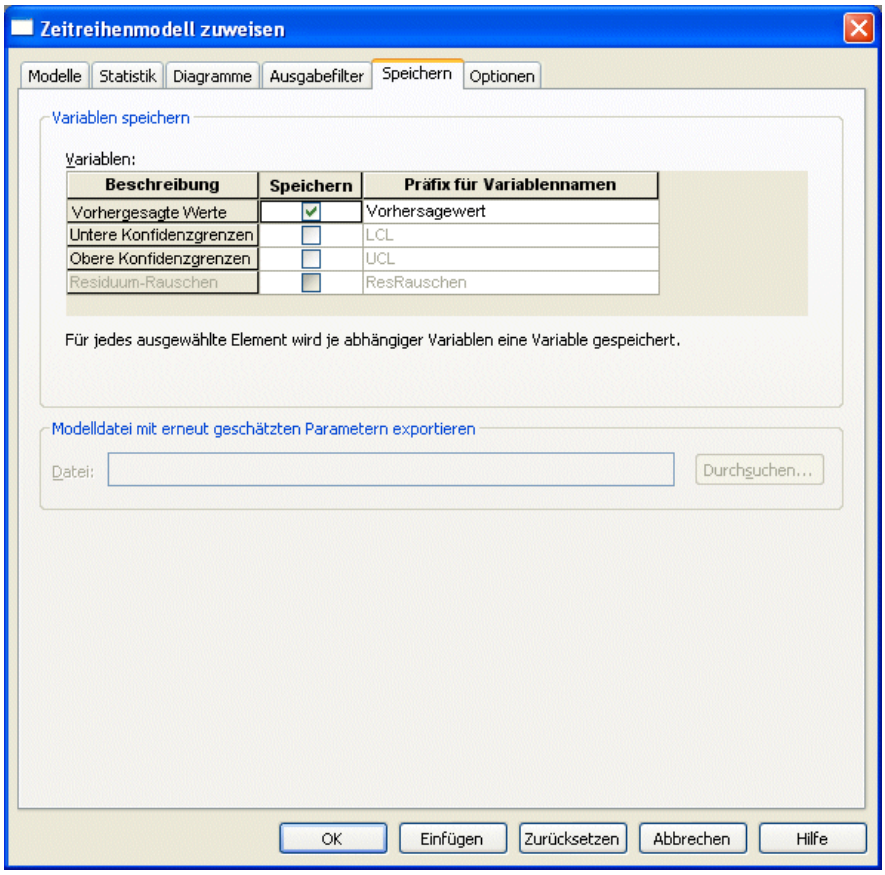
- Wählen Sie im Gruppenfeld "Vorhersageperiode" die Option Erster Fall nach der Schätzperiode bis zum angegebenen Datum aus.

▶ Geben Sie im Datumsgitter für das Jahr 2004 und für den Monat 6 ein.

Das Daten-Set enthält Daten von Januar 1999 bis März 2004. Mit den aktuellen Einstellungen erstreckt sich die Vorhersageperiode von April 2004 bis Juni 2004.

▶ Klicken Sie auf die Registerkarte Speichern.

Abbildung 7-2
Zeitreihenmodell zuweisen, Registerkarte "Speichern"

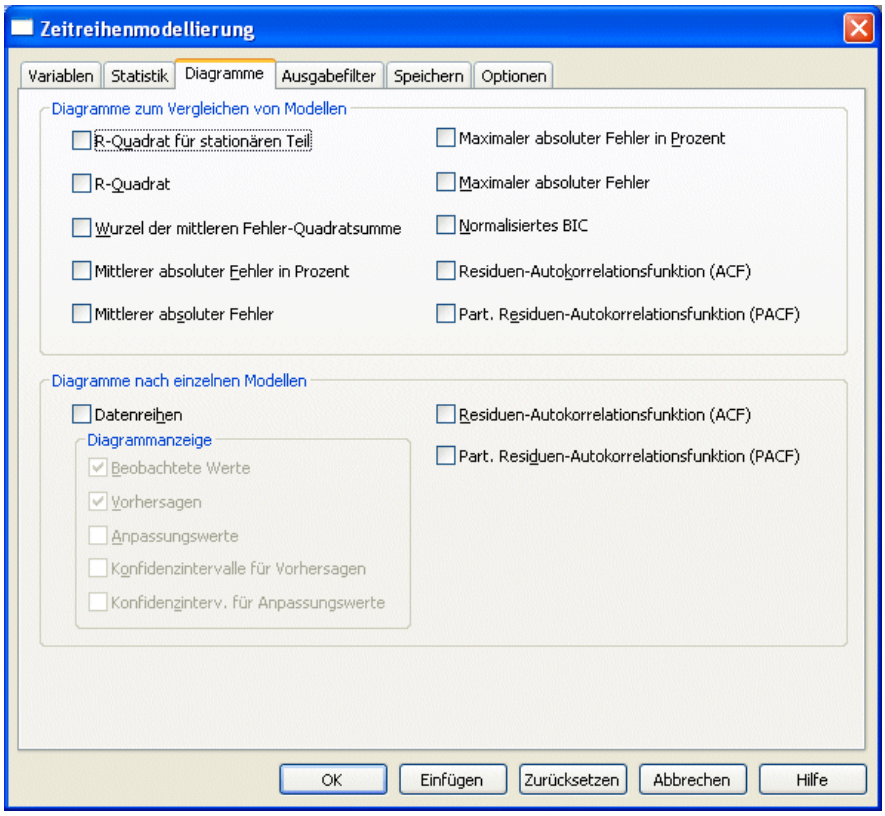


▶ Aktivieren Sie in der Spalte *Speichern* den Eintrag "Vorhergesagte Werte", und behalten Sie den Standardwert *Vorhersagewert* als "Präfix für Variablennamen" bei.

Die Modellvorhersagen werden als neue Variablen in der Arbeitsdatei gespeichert, wobei *Vorhersagewert* als Präfix für die Variablennamen verwendet wird.

▶ Klicken Sie auf die Registerkarte Diagramme.

Abbildung 7-3
Zeitreihenmodell zuweisen, Registerkarte "Diagramme"



▶ Deaktivieren Sie im Gruppenfeld "Diagramme nach einzelnen Modellen" die Option Datenreihen.

Hierbei werden die Zeitreihen-Diagramme für einzelne Modelle unterdrückt. Das vorliegende Beispiel soll in erster Linie erläutern, wie die Vorhersagen in neuen Variablen gespeichert werden. Das Erzeugen von Diagrammen der Vorhersagen ist nur von sekundärem Interesse.

▶ Klicken Sie im Dialogfeld "Zeitreihenmodell zuweisen" auf OK.

Statistiken zur Anpassungsgüte des Modells

Abbildung 7-4
Anpassungsgüte des Modells (Tabelle)

Anpassungsstatistik	Mittelwert	Standardfehler	Minimum	Maximum	Perzentil 5	10	25	50	75	90	95
R-Quadrat für stationären Teil	,313	,178	,007	,610	,062	,094	,170	,266	,494	,577	,590
R-Quadrat	,999	,000	,998	1,000	,999	,999	,999	,999	,999	,999	1,000
RMSE	186,630	150,840	44,186	812,212	52,315	56,009	90,830	138,619	197,176	426,084	482,442
MAPE	,858	,081	,667	1,136	,733	,759	,804	,854	,916	,964	,977
MaxAPE	2,415	,500	1,634	4,164	1,836	1,898	2,072	2,309	2,593	2,952	3,718
MAE	143,024	112,632	34,555	632,596	40,643	45,432	70,103	107,809	154,517	321,606	365,637
MaxAE	497,210	424,651	114,927	2269,688	124,157	139,799	222,208	368,380	567,763	1182,352	1358,681
Normalisiertes BIC	10,116	1,352	7,774	13,531	8,095	8,248	9,215	9,995	10,732	12,240	12,489

Die Tabelle “Anpassungsgüte des Modells” enthält modellübergreifend berechnete Statistiken zur Anpassungsgüte. Diese Tabelle ist eine präzise Auswertung, wie gut das Modell mit den neu geschätzten Parametern an die Daten angepasst ist. Für jede Statistik enthält die Tabelle modellübergreifend den Mittelwert, den Standardfehler, den Minimal- und den Maximalwert. Darüber hinaus enthält sie Perzentilwerte für die Verteilung der Statistik über die Modelle. Das jeweilige Perzentil gibt den Prozentsatz der Modelle an, die einen Wert der Anpassungsstatistik aufweisen, der unter dem angegebenen Wert liegt. Beispielsweise weisen 95 % der Modelle einen maximalen absoluten Fehler in Prozent (MaxAPE) unter 3,676 auf.

Es werden mehrere Statistiken ausgegeben, hier sind jedoch zwei von besonderer Bedeutung: MAPE (mittlerer absoluter Fehler in Prozent) und MaxAPE (maximaler absoluter Fehler in Prozent). Der absolute Fehler in Prozent ist ein Maß dafür, wie sehr eine abhängige Zeitreihe vom Niveau abweicht, das vom Modell vorhergesagt wurde, und bietet einen Hinweis auf die Unsicherheit der Vorhersage. Der mittlere absolute Fehler in Prozent variiert über alle Modelle zwischen einem Minimum von 0,669 % und einem Maximum von 1,026 %. Der maximale absolute Fehler in Prozent variiert über alle Modelle zwischen 1,742 % und 4,373 %. Die mittlere Unsicherheit in den Vorhersagen der einzelnen Modelle beträgt daher etwa 1 % und die maximale Unsicherheit liegt bei etwa 2,5 % (dem Mittelwert von MaxAPE), wobei der schlechteste anzunehmende Fall bei etwa 4 % liegt. Ob diese Werte einer akzeptablen Unsicherheit entsprechen, ist von dem Risiko abhängig, das Sie zu akzeptieren bereit sind.

Modellvorhersagen

Abbildung 7-5
Neue Variablen mit Modellvorhersagen

YEAR_	MONTH_	DATE_	Vorhersagewert_Market_1_Modell_1	Vorhersagewert_Market_2_2 Modell_2
2004	1	JAN 2004	11513	54947
2004	2	FEB 2004	11806	56810
2004	3	MAR 2004	11950	57344
2004	4	APR 2004	12312	59631
2004	5	MAY 2004	12501	60717
2004	6	JUN 2004	12689	61659

Der Daten-Editor zeigt die neuen Variablen an, die die Modellvorhersagen enthalten. Auch wenn hier nur zwei gezeigt werden, sind 85 neue Variablen vorhanden: eine für jede der 85 abhängigen Zeitreihen. Die Variablennamen bestehen aus dem Standardpräfix *Vorhersagewert*, gefolgt vom Namen der verknüpften abhängigen Variable (zum Beispiel *Market_1* und einem Modellbezeichner (zum Beispiel *Modell_1*).

Dem Daten-Set wurden drei neue Fälle mit den Vorhersagen von April 2004 bis Juni 2005 und automatisch erzeugten Datenlabels hinzugefügt.

Zusammenfassung

Sie haben gelernt, wie gespeicherte Modelle zugewiesen werden, um frühere Vorhersagen bei der Verfügbarkeit von neuen Daten zu erweitern. Dies haben Sie ohne erneutes Erstellen der Modelle erledigt. Wenn ein Grund zur Annahme besteht, dass ein Modell sich geändert hat, sollten Sie es selbstverständlich mit der Prozedur “Zeitreihenmodellierung” neu erstellen.

Kapitel
8

Ermitteln von signifikanten Einflussvariablen mit dem Expert Modeler

Eine Katalogfirma, die an der Entwicklung eines Prognosemodells interessiert ist, hat Daten zum monatlichen Absatz von Herrenbekleidung sowie mehrere Zeitreihen gesammelt, die verwendet werden können, um einen Teil der Absatzschwankungen zu erklären. Zu den möglichen Einflussvariablen gehören die Anzahl der versendeten Kataloge, die Anzahl der Seiten im Katalog, die Anzahl der Telefonleitungen, über die eine Bestellung möglich ist, die Ausgaben für Werbung in Printmedien und die Anzahl der Kundendienstmitarbeiter. Sind diese Einflussvariablen sinnvoll für die Prognostizierung?

In diesem Beispiel ermitteln Sie mit dem Expert Modeler in Verbindung mit allen potenziellen Einflussvariablen das beste Modell. Da der Expert Modeler nur die Einflussvariablen auswählt, die eine statistisch signifikante Beziehung zu den abhängigen Datenreihen aufweisen, erfahren Sie so, welche Einflussvariablen nützlich sind, und erhalten ein diesbezügliches Modell für Vorhersagen. Wenn Sie fertig sind, empfiehlt es sich möglicherweise, das nächste Beispiel, Untersuchen von Einflussvariablen durch das Zuweisen gespeicherter Modelle in Kapitel 9 auf S. 77, durchzuarbeiten, in dem die Auswirkungen verschiedener Szenarien mit Einflussvariablen auf Verkäufe mithilfe des in diesem Beispiel erstellten Modells untersucht werden.

Die Daten für das aktuelle Beispiel befinden sich in *catalog_seasfac.sav*. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108.

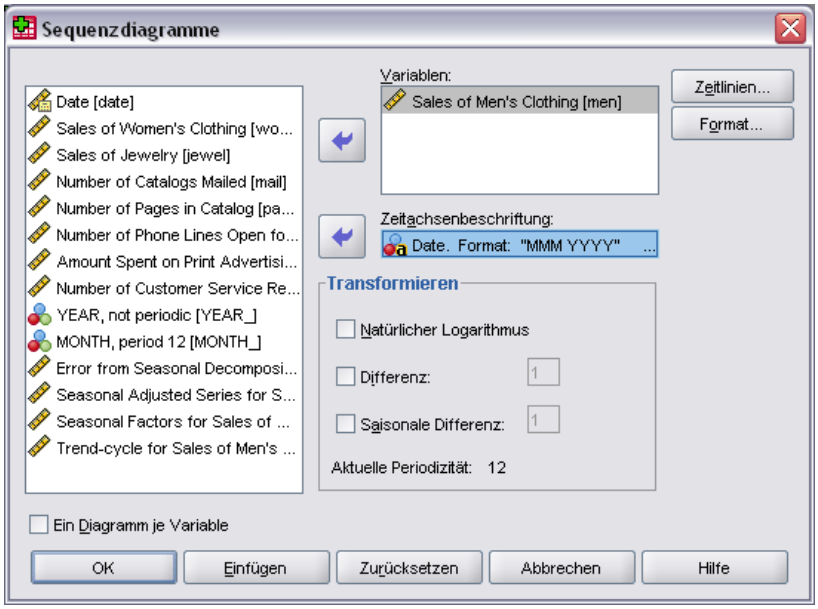
Erstellen von Diagrammen aus den Daten

Besonders bei der Arbeit mit nur einer Datenreihe sollten Sie die Daten immer grafisch darstellen.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

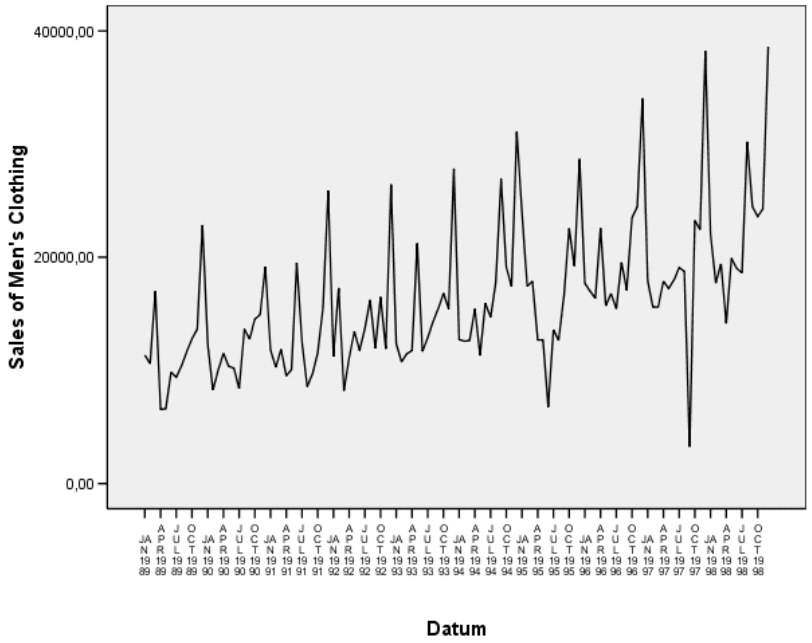
Analysieren
  Zeitreihen
    Sequenzdiagramme...

Abbildung 8-1
Dialogfeld "Sequenzdiagramme"



- Wählen Sie die Variable *Sales of Men's Clothing* aus und verschieben Sie sie in die Liste "Variablen".
- Wählen Sie die Variable *Date* aus und verschieben Sie sie in das Feld "Zeitachsenbeschriftung".
- Klicken Sie auf OK.

Abbildung 8-2
Absatz an Herrenbekleidung (in US-Dollar)



Die Datenreihe weist zahlreiche Spitzen auf, die häufig scheinbar in gleichen Abständen zu finden sind, und zeigt einen deutlichen Aufwärtstrend. Diese Spitzen mit gleichen Abständen deuten auf das Vorhandensein einer periodischen Komponente in der Zeitreihe hin. In Anbetracht der saisonalen Natur der Verkäufe, die normalerweise in der Weihnachtszeit Spitzenwerte aufweisen, dürften Sie nicht verwundert sein, eine jährliche saisonale Komponente in den Daten zu finden.

Außerdem sind Spitzen vorhanden, die nicht als Teil des Saisonmusters erscheinen und signifikante Abweichungen von den benachbarten Datenpunkten darstellen. Bei diesen Punkten kann es sich um Ausreißer handeln, die mit dem Expert Modeler bearbeitet werden können und sollten.

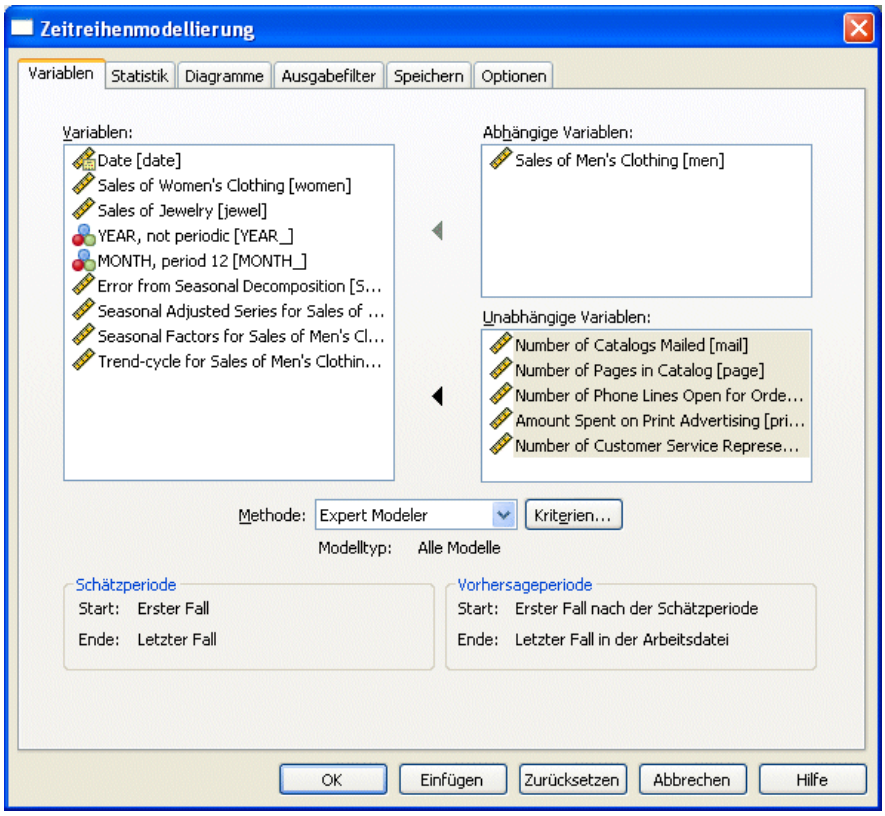
Durchführung der Analyse

So verwenden Sie den Expert Modeler:

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

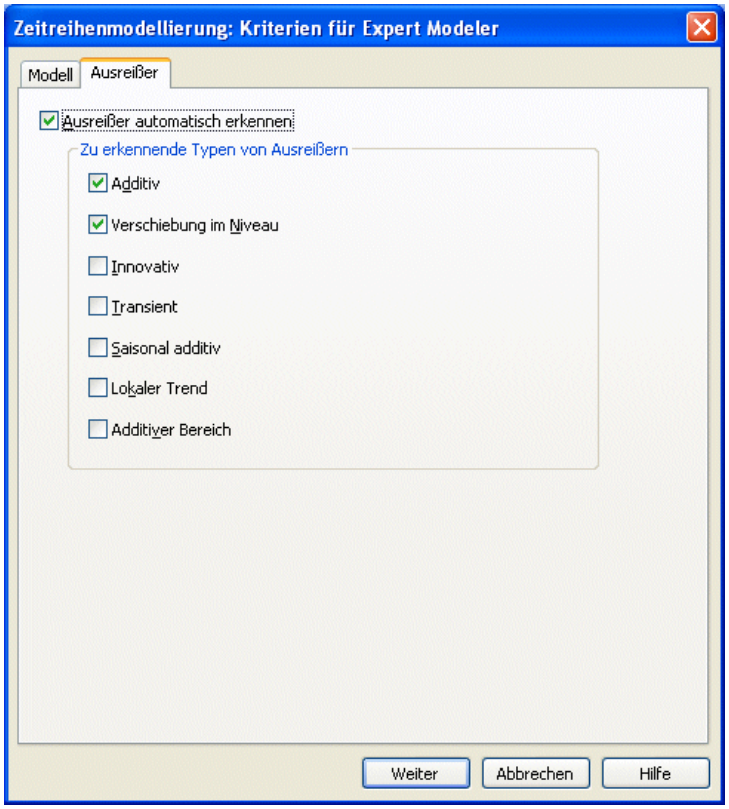
Analysieren
  Zeitreihen
    Modelle erstellen...

Abbildung 8-3
Dialogfeld "Zeitreihenmodellierung"



- Wählen Sie *Sales of Men's Clothing* als abhängige Variable aus.
- Wählen Sie die Variablen *Number of Catalogs Mailed* bis *Number of Customer Service Representatives* als unabhängige Variablen aus.
- Stellen Sie sicher, dass in der Dropdown-Liste "Methode" die Option Expert Modeler ausgewählt ist. Der Expert Modeler ermittelt automatisch das jeweils am besten angepasste saisonale oder nichtsaisonale Modell für jede der abhängigen Zeitreihen.
- Klicken Sie auf Kriterien, und klicken Sie dann auf die Registerkarte Ausreißer.

Abbildung 8-4
Dialogfeld "Kriterien für Expert Modeler", Registerkarte "Ausreißer"



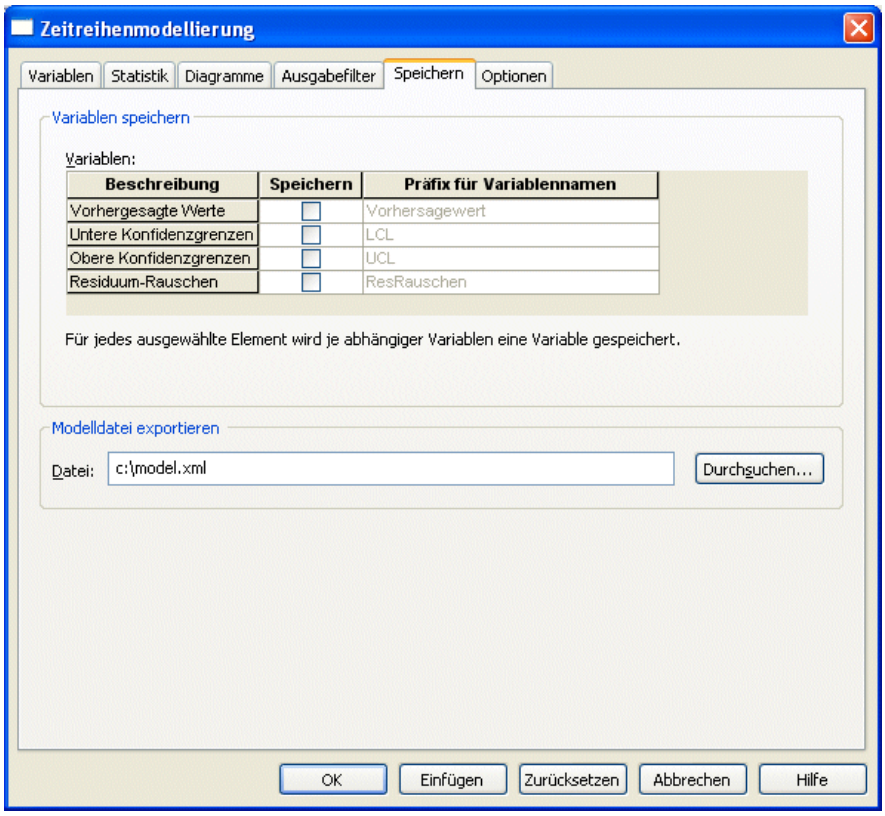
- Wählen Sie Ausreißer automatisch erkennen, und behalten Sie die Standardauswahl für die zu erkennenden Ausreißertypen unverändert bei.

  Bei der visuellen Untersuchung der Daten stellte sich heraus, dass möglicherweise Ausreißer vorliegen. Mit der aktuellen Auswahl sucht der Expert Modeler nach den häufigsten Ausreißertypen und integriert etwaige Ausreißer in das endgültige Modell. Die Ausreißererkennung kann zu beträchtlichen Erhöhungen der erforderlichen Rechenzeit für den Expert Modeler führen. Daher sollte diese Funktion mit Bedacht verwendet werden, insbesondere wenn Sie viele Datenreihen gleichzeitig modellieren. In der Standardeinstellung werden Ausreißer nicht erkannt.

- Klicken Sie auf Weiter.

- Klicken Sie im Dialogfeld "Zeitreihenmodellierung" auf die Registerkarte Speichern.

Abbildung 8-5
Zeitreihenmodellierung, Registerkarte "Speichern"



Das geschätzte Modell sollte in einer externen XML-Datei gespeichert werden, sodass Sie (in der Prozedur "Zeitreihenmodell zuweisen") mit verschiedene Werte für die Einflussvariablen experimentieren können, ohne das Modell erneut erstellen zu müssen.

- Klicken Sie auf der Registerkarte "Speichern" auf die Schaltfläche Durchsuchen.

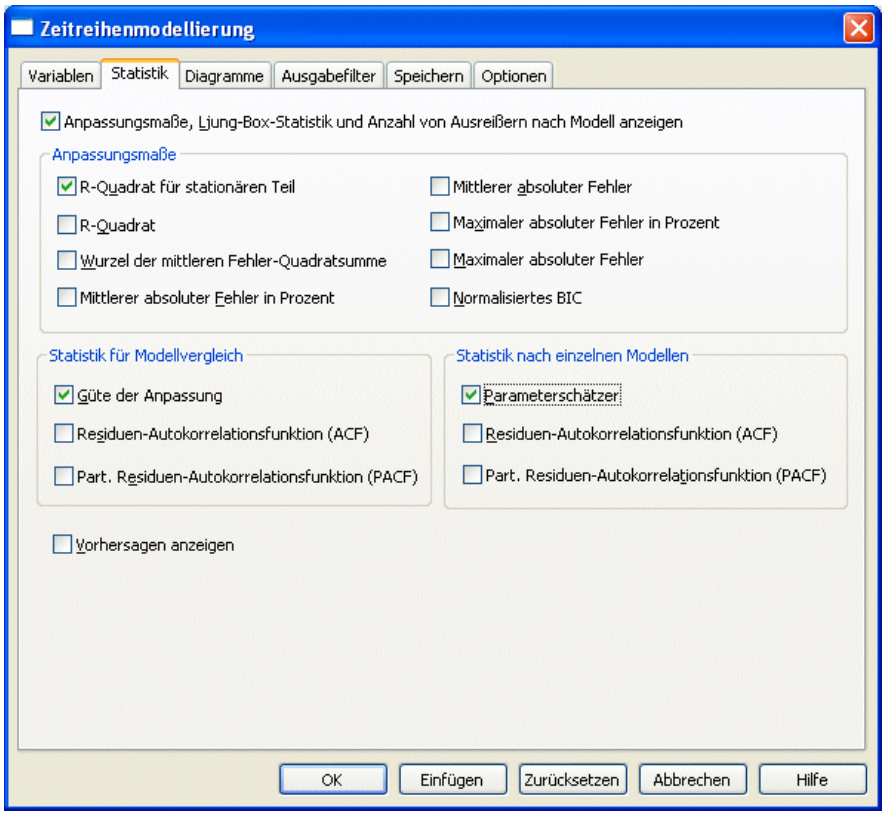
  Dadurch wird ein Standarddialogfeld zum Speichern von Dateien geöffnet.

- Wechseln Sie in den Ordner, in dem Sie die XML-Modelldatei speichern möchten. Geben Sie einen Dateinamen ein, und klicken Sie auf Speichern.

  Der Pfad für die XML-Modelldatei wird daraufhin auf der Registerkarte "Speichern" angezeigt.

- Klicken Sie auf die Registerkarte Statistik.

Abbildung 8-6
Zeitreihenmodellierung, Registerkarte "Statistik"

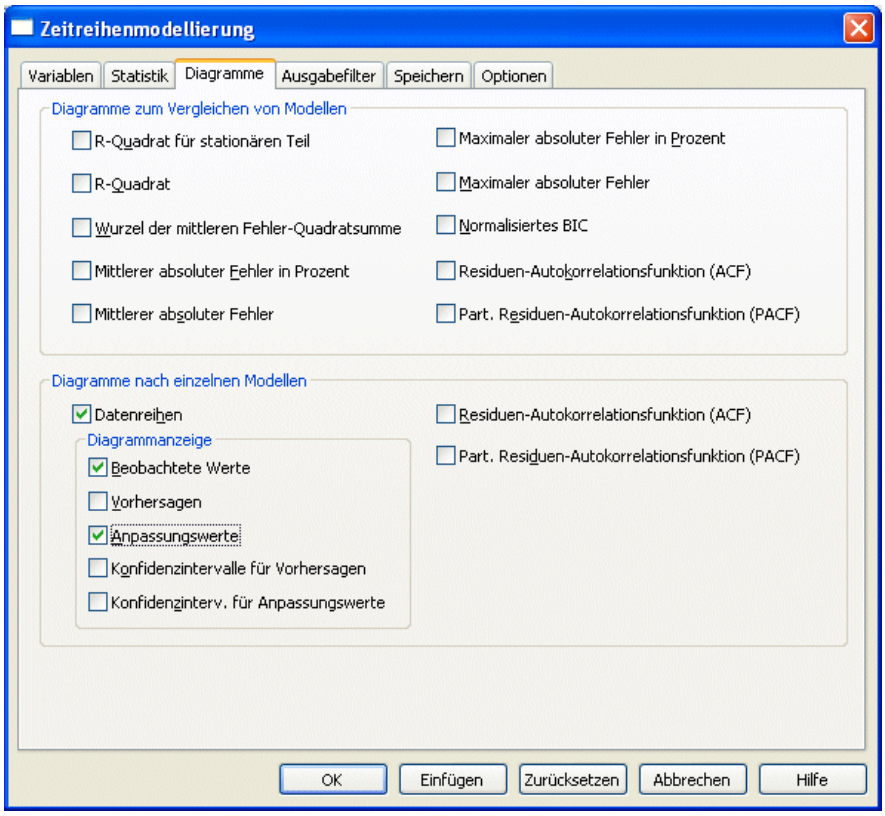


- Wählen Sie Parameterschätzer aus.

  Mit dieser Option wird eine Tabelle erstellt, die alle Parameter für das vom Expert Modeler gewählte Modell einschließlich der signifikanten Einflussvariablen enthält.

- Klicken Sie auf die Registerkarte Diagramme.

Abbildung 8-7
Zeitreihenmodellierung, Registerkarte "Diagramme"



▶ Deaktivieren Sie die Option Vorhersagen.

Das vorliegende Beispiel soll nur erläutern, wie signifikante Einflussvariablen ermittelt und ein Modell erstellt wird. Es werden keine Vorhersagen aufgestellt.

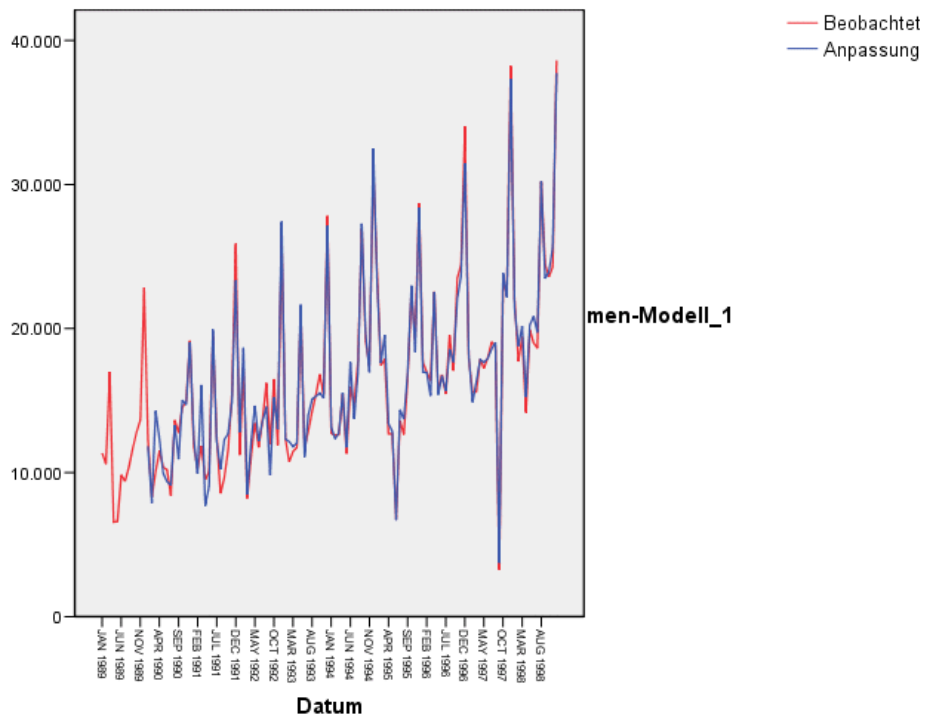
▶ Wählen Sie Anpassungswerte aus.

Bei dieser Option werden die vorhergesagten Werte für die Periode angezeigt, in der das Modell geschätzt wird. Diese Periode wird als **Schätzperiode** bezeichnet und enthält alle Fälle in der Arbeitsdatei für dieses Beispiel. Diese Werte bieten einen Hinweis darauf, wie gut das Modell an die beobachteten Werte angepasst ist. Daher werden sie als **Anpassungswerte** bezeichnet. Das resultierende Diagramm enthält die beobachteten Werte und die Anpassungswerte.

▶ Klicken Sie im Dialogfeld "Zeitreihenmodellierung" auf OK.

Zeitreihendiagramm

Abbildung 8-8
Vorhergesagte und beobachtete Werte



Die vorhergesagten Werte entsprechen den beobachteten Werten recht gut. Dies deutet darauf hin, dass das Modell eine zufrieden stellende Vorhersagekraft besitzt. Beachten Sie, wie genau die saisonalen Spitzen im Modell vorhergesagt werden. Der Aufwärtstrend der Daten wird ebenfalls angemessen erfasst.

Tabelle der Modellbeschreibung

Abbildung 8-9
Tabelle der Modellbeschreibung

			Modelltyp
Modell-ID	Sales of Men's Clothing	Modell_1	ARIMA(0,0,0)(0,1,0)

Die Tabelle der Modellbeschreibung enthält einen Eintrag für jedes geschätzte Modell mit Modellbezeichner und Modelltyp. Der Modellbezeichner umfasst den Namen (oder das Label) der zugeordneten abhängigen Variablen sowie einen vom System zugewiesenen Namen. Im vorliegenden Beispiel ist die abhängige Variable *Sales of Men's Clothing*, und der vom System zugewiesene Name lautet *Modell_1*.

Die Zeitreihenmodellierung unterstützt sowohl Modelle mit exponentiellem Glätten als auch ARIMA-Modelle. Die Modelltypen mit exponentiellem Glätten werden nach den allgemein verwendeten Namen aufgeführt, z. B. “Holt” oder “Winters-additiv”. ARIMA-Modelltypen werden nach der Standardnotation für ARIMA aufgeführt (p, d, q)(P, D, Q), wobei p die Ordnung

der Autoregression, *d* die Ordnung der Differenzbildung (oder Integration) und *q* die Ordnung des gleitenden Durchschnitts ist und (*P*, *D*, *Q*) deren saisonalen Entsprechungen darstellen.

Der Expert Modeler hat ermittelt, dass der Verkauf von Herrenbekleidung am besten durch ein saisonales ARIMA-Modell mit Differenzbildung erster Ordnung beschrieben wird. Mit der saisonalen Natur des Modells werden die saisonalen Spitzen berücksichtigt, die im Zeitreihendiagramm zu erkennen waren, und die Differenzbildung der ersten Ordnung entspricht dem in den Daten abzulesenden Aufwärtstrend.

Tabelle der Modellstatistik

Abbildung 8-10
Tabelle der Modellstatistik

Modell	Anzahl der Prädiktoren	Statistiken zur Güte der Anpassung	Ljung-Box-Q(18)			Anzahl der Ausreißer
		R-Quadrat für stationären Teil	Statistiken	Freiheitsgrade	Sig.	
Sales of Men's Clothing-Modell_1	2	,948	7,589	18	,984	9

Die Tabelle der Modellstatistik enthält Auswertungsinformationen und Anpassungsstatistiken für jedes geschätzte Modell. Die Ergebnisse jedes Modells sind mit dem Modellbezeichner aus der Tabelle der Modellbeschreibung beschriftet. Beachten Sie zunächst, dass das Modell von den ursprünglich angegebenen fünf potenziellen Einflussvariablen zwei enthält. Der Expert Modeler hat offenbar zwei unabhängige Variablen ermittelt, die für die Vorhersage sinnvoll sein könnten.

Obwohl die Zeitreihenmodellierung verschiedene Anpassungsstatistiken bietet, wurde hier lediglich *R*-Quadrat für den stationären Teil gewählt. Diese Statistik bietet eine Schätzung des Anteils der Gesamtvariation in der Datenreihe, der durch das Modell erklärt wird. Sie ist gegenüber dem normalen *R*-Quadrat vorzuziehen, wenn wie im vorliegenden Beispiel ein Trend oder ein saisonales Muster vorliegt. Größere Werte von *R*-Quadrat für den stationären Teil (bis zu einem Maximalwert von 1) weisen auf eine bessere Anpassung hin. Ein Wert von 0,948 bedeutet, dass das Modell für die Erklärung der beobachteten Variation in der Datenreihe bestens geeignet ist.

Die Box-Ljung-Statistik, auch “modifizierte Box-Pierce-Statistik” genannt, bietet einen Hinweis darauf, ob das Modell ordnungsgemäß spezifiziert wurde. Ein Signifikanzwert unter 0,05 weist darauf hin, dass in der beobachteten Datenreihe Strukturen vorhanden sind, die im Modell nicht berücksichtigt werden. Der hier gezeigte Wert 0,984 ist nicht signifikant. Daher kann davon ausgegangen werden, dass das Modell richtig spezifiziert wurde.

Der Expert Modeler hat neun Punkte als Ausreißer erkannt. Jeder dieser Punkte wurde ordnungsgemäß modelliert, daher müssen sie nicht aus der Datenreihe entfernt werden.

Tabelle der ARIMA-Modellparameter

Abbildung 8-11
Tabelle der ARIMA-Modellparameter

					Schätzer	Standardfehler	t	Sig.
Sales of Men's Clothing-Modell_1	Sales of Men's Clothing	Keine Transformation	Saisonale Differenz		1			
	Number of Catalogs Mailed	Keine Transformation	Zähler	Lag 0	1,549	,071	21,943	,000
			Saisonale Differenz		1			
	Number of Phone Lines Open for Ordering	Keine Transformation	Zähler	Lag 0	315,262	15,298	20,607	,000
			Saisonale Differenz		1			

In der Tabelle der ARIMA-Modellparameter werden Werte für alle Parameter im Modell aufgeführt, wobei für jedes geschätzte Modell ein mit dem Modellbezeichner beschrifteter Eintrag vorhanden ist. Im Rahmen des Beispiels werden alle Variablen im Modell aufgeführt, einschließlich der abhängigen Variablen und allen unabhängigen Variablen, die vom Expert Modeler als signifikant eingestuft wurden. Aus der Tabelle der Modellstatistik ist bereits bekannt, dass zwei signifikante Einflussvariablen vorhanden sind. Die Tabelle der Modellparameter zeigt, dass es sich um *Number of Catalogs Mailed* und *Number of Phone Lines Open for Ordering* handelt.

Zusammenfassung

Sie haben gelernt, wie Sie mit dem Expert Modeler ein Modell erstellen und signifikante Einflussvariablen ermitteln, und Sie haben das resultierende Modell als externe Datei gespeichert. Sie können jetzt mit der Prozedur “Zeitreihenmodell zuweisen” alternative Szenarien für die Prädiktorzeitreihe untersuchen und bewerten, wie sich die Alternativen auf die Absatzvorhersagen auswirken.

Kapitel
9

Untersuchen von Einflussvariablen durch das Zuweisen gespeicherter Modelle

Sie haben mit der Zeitreihenmodellierung ein Modell Ihrer Daten erstellt und die für Vorhersagen geeigneten Einflussvariablen identifiziert. Die Einflussvariablen stellen Faktoren dar, die Sie beeinflussen können. Daher möchten Sie unterschiedliche Werte für diese Variablen in der Vorhersageperiode untersuchen, um deren Einfluss auf die abhängige Variable zu ermitteln. Diese Aufgabe kann problemlos mit der Prozedur "Zeitreihenmodell zuweisen" durchgeführt werden, indem Sie die in der Prozedur "Zeitreihenmodellierung" erstellte Modelldatei verwenden.

Dieses Beispiel ist eine Erweiterung des vorhergehenden Beispiels, Ermitteln von signifikanten Einflussvariablen mit dem Expert Modeler in Kapitel 8 auf S. 66, kann jedoch auch unabhängig verwendet werden. Das Szenario betrifft ein Katalogunternehmen, das Daten über den monatlichen Umsatz von Herrenbekleidung von Januar 1989 bis Dezember 1998 sowie weitere Datenreihen erfasst hat, die als potenziell nützliche Einflussvariablen für den zukünftigen Umsatz angesehen werden. Mit dem Expert Modeler wurde ermittelt, dass nur zwei der fünf potenziellen Einflussvariablen signifikant sind: die Anzahl der per Post versendeten Kataloge und die Anzahl der für Bestellungen verfügbaren Telefonleitungen.

Für die Vertriebsstrategie im nächsten Jahr verfügen Sie über begrenzte Ressourcen zum Drucken von Katalogen und zum Bereitstellen verfügbarer Telefonleitungen für die Bestellungsannahme. Ihr Budget für die ersten drei Monate 1999 ermöglicht über die ursprünglichen Prognosen hinaus entweder 2000 zusätzliche Kataloge oder 5 zusätzliche Telefonleitungen. Welche der beiden Möglichkeiten führt in diesem Zeitraum von drei Monaten zu mehr Umsatzerlösen?

Die Daten für das Modell befinden sich in *catalog_seasfac.sav*; *catalog_model.xml* enthält das mit dem Expert Modeler erstellte Modell des monatlichen Umsatzes. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108. Wenn Sie das vorhergehende Beispiel durchgearbeitet und eine eigene Modelldatei gespeichert haben, können Sie selbstverständlich anstelle von *catalog_model.xml* Ihre eigene Datei verwenden.

Erweitern der Prädiktorzeitreihe

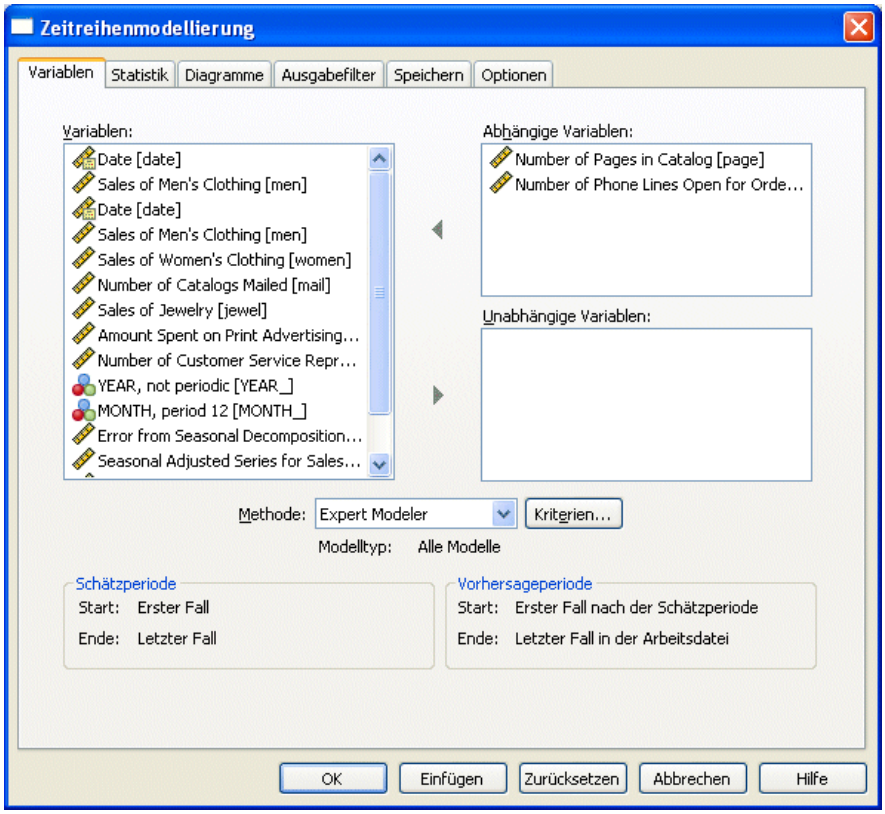
Beim Erstellen von Vorhersagen für abhängige Datenreihen mit Einflussvariablen muss jede Prädiktorzeitreihe über die Vorhersageperiode hinaus erweitert werden. Wenn Ihnen die zukünftigen Werte der Einflussvariablen nicht genau bekannt sind, müssen Sie diese schätzen. Anschließend können Sie die Schätzwerte ändern und so verschiedene Szenarien mit

Einflussvariablen testen. Die anfänglichen Prognosen können einfach mit dem Expert Modeler erstellt werden.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Modelle erstellen...

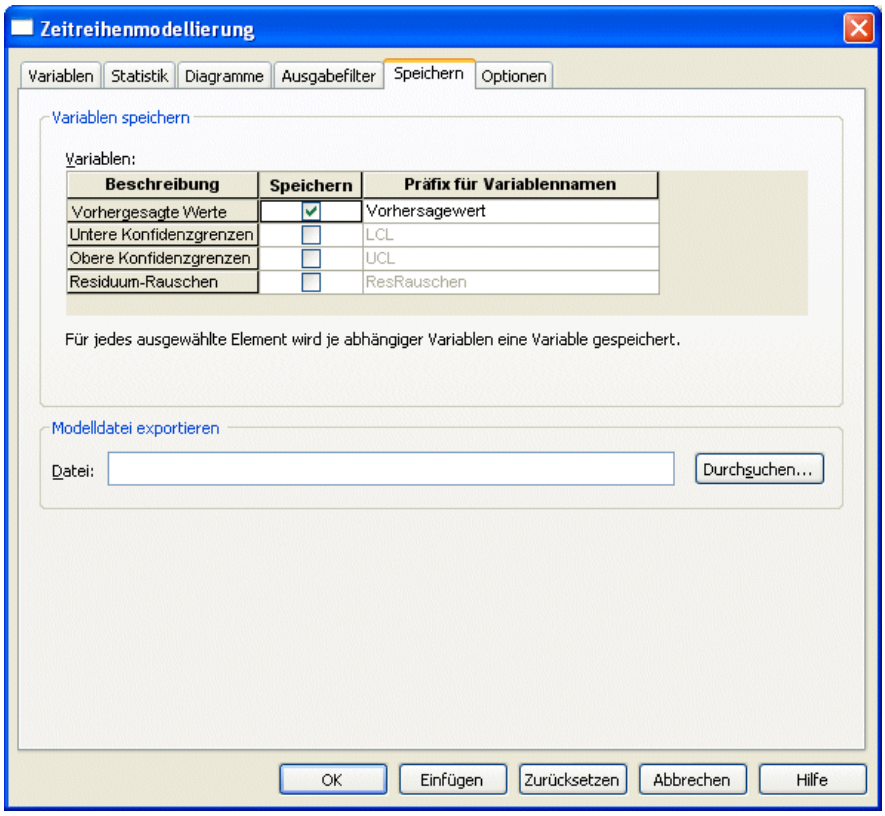
Abbildung 9-1
Dialogfeld "Zeitreihenmodellierung"



▶ Wählen Sie *Number of Catalogs Mailed* und *Number of Phone Lines Open for Ordering* als abhängige Variablen aus.

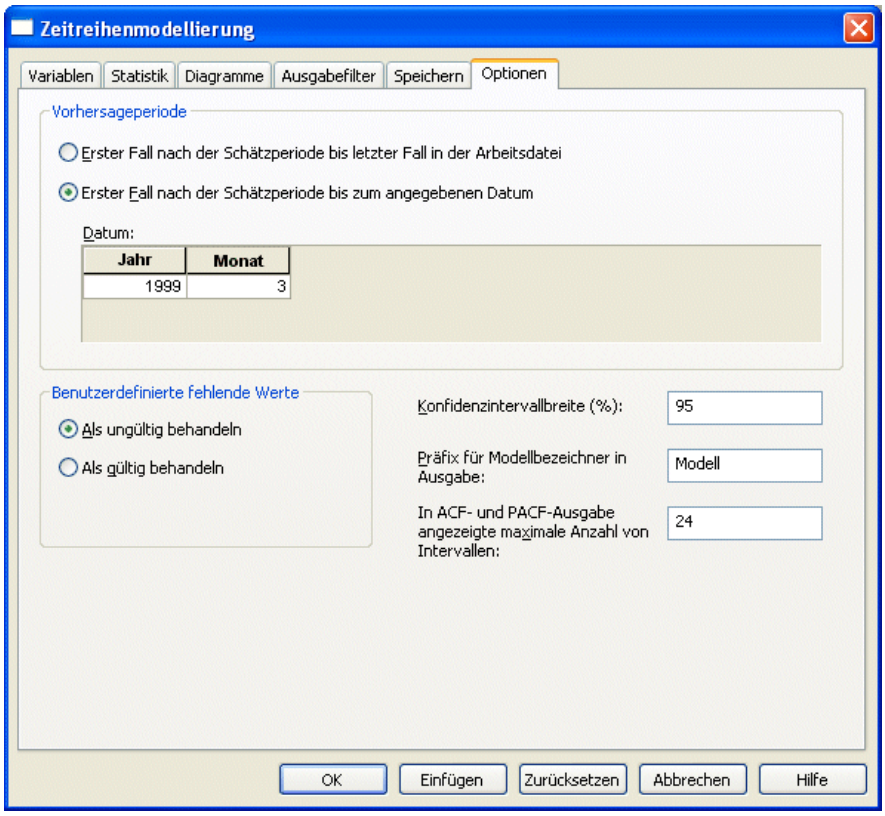
▶ Klicken Sie auf die Registerkarte Speichern.

Abbildung 9-2
Zeitreihenmodellierung, Registerkarte "Speichern"



- Wählen Sie in der Spalte *Speichern* den Eintrag "Vorhergesagte Werte" aus und behalten Sie den Standardwert *Vorhersagewert* als "Präfix für Variablennamen" bei.
- Klicken Sie auf die Registerkarte Optionen.

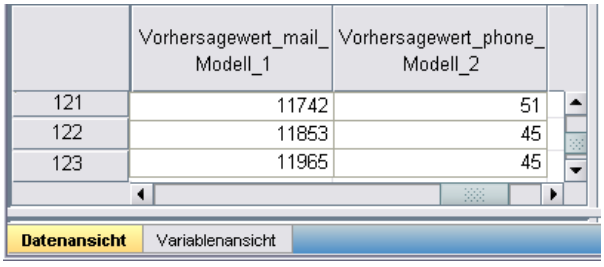
Abbildung 9-3
Zeitreihenmodellierung, Registerkarte "Optionen"



- Wählen Sie im Gruppenfeld "Vorhersageperiode" die Option Erster Fall nach der Schätzperiode bis zum angegebenen Datum aus.
- Geben Sie im Datumsgitter für das Jahr 1999 und für den Monat 3 ein.

  Das Daten-Set enthält Daten von Januar 1989 bis Dezember 1998, sodass mit den aktuellen Einstellungen die Vorhersageperiode Januar 1999 bis März 1999 umfasst.
- Klicken Sie auf OK.

Abbildung 9-4
Neue Variablen mit Vorhersagen für Prädiktorzeitreihen



Im Daten-Editor werden die neuen Variablen *Vorhersagewert_mail_Modell_1* und *Vorhersagewert_phone_Modell_2* angezeigt, die die durch das Modell vorhergesagten Werte für die Anzahl der per Post versendeten Kataloge und die Anzahl der Telefonleitungen enthalten.

Zum Erweitern der Prädiktorzeitreihe werden nur Werte für Januar 1999 bis März 1999 benötigt, also die Fälle 121 bis 123.

- Kopieren Sie die Werte dieser drei Fälle aus *Vorhersagewert_mail_Modell_1*, und hängen Sie sie an die Variable *mail* an.
- Wiederholen Sie diesen Vorgang für *Vorhersagewert_phone_Modell_2*, wobei Sie die letzten drei Fälle kopieren und an die Variable *phone* anhängen.

Abbildung 9-5
Über die Vorhersageperiode hinaus erweiterte Prädiktorzeitreihe

	mail	page	phone	print	service	YEAR_	MONTH_	DATE_
121	11742	.	51	.	.	1999	1	JAN 1999
122	11853	.	45	.	.	1999	2	FEB 1999
123	11965	.	45	.	.	1999	3	MAR 1999

Datenansicht | Variablenansicht

Die Einflussvariablen wurden nun über die Vorhersageperiode hinaus erweitert.

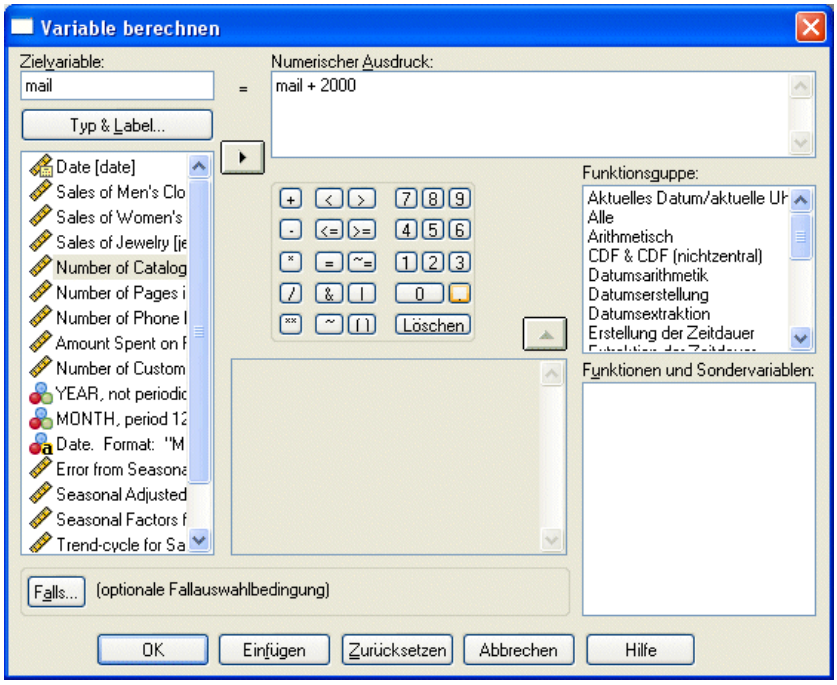
Ändern der Werte von Einflussvariablen in der Vorhersageperiode

Zum Testen der beiden Szenarien (Versenden zusätzlicher Kataloge oder Bereitstellen zusätzlicher Telefonleitungen) müssen die Schätzwerte für die Einflussvariablen *mail* oder *phone* geändert werden. Da hier nur die Einflussvariablen für drei Fälle (Monate) geändert werden, könnten die neue Werte einfach direkt in den entsprechenden Zellen im Daten-Editor eingegeben werden. Um das entsprechende Verfahren zu erläutern, wird jedoch das Dialogfeld “Variable berechnen” verwendet. Wenn viele Werte zu ändern sind, ist das Dialogfeld “Variable berechnen” meist vorzuziehen.

- Wählen Sie die folgenden Befehle aus den Menüs aus:

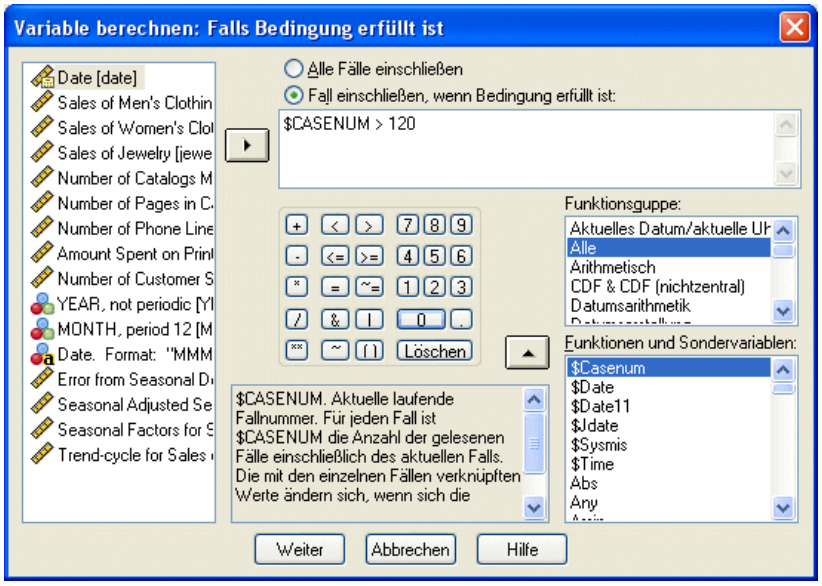
Transformieren
  Variable berechnen...

Abbildung 9-6
Dialogfeld "Variable berechnen"



- Geben Sie mail als Zielvariable ein.
- Geben Sie im Textfeld "Numerischer Ausdruck" die Zeichenfolge mail + 2000 ein.
- Klicken Sie auf Falls.

Abbildung 9-7
Dialogfeld "Variable berechnen: Falls Bedingung erfüllt ist"



- Wählen Sie Fall einschließen, wenn Bedingung erfüllt ist aus.
- Geben Sie im Textfeld die Zeichenfolge $CASENUM > 120 ein.

  Dadurch werden Änderungen an der Variablen *mail* auf die Fälle in der Vorhersageperiode begrenzt.
- Klicken Sie auf Weiter.
- Klicken Sie im Dialogfeld "Variable berechnen" auf OK, und klicken Sie erneut auf OK, wenn Sie gefragt werden, die Sie die vorhandene Variable ändern möchten.

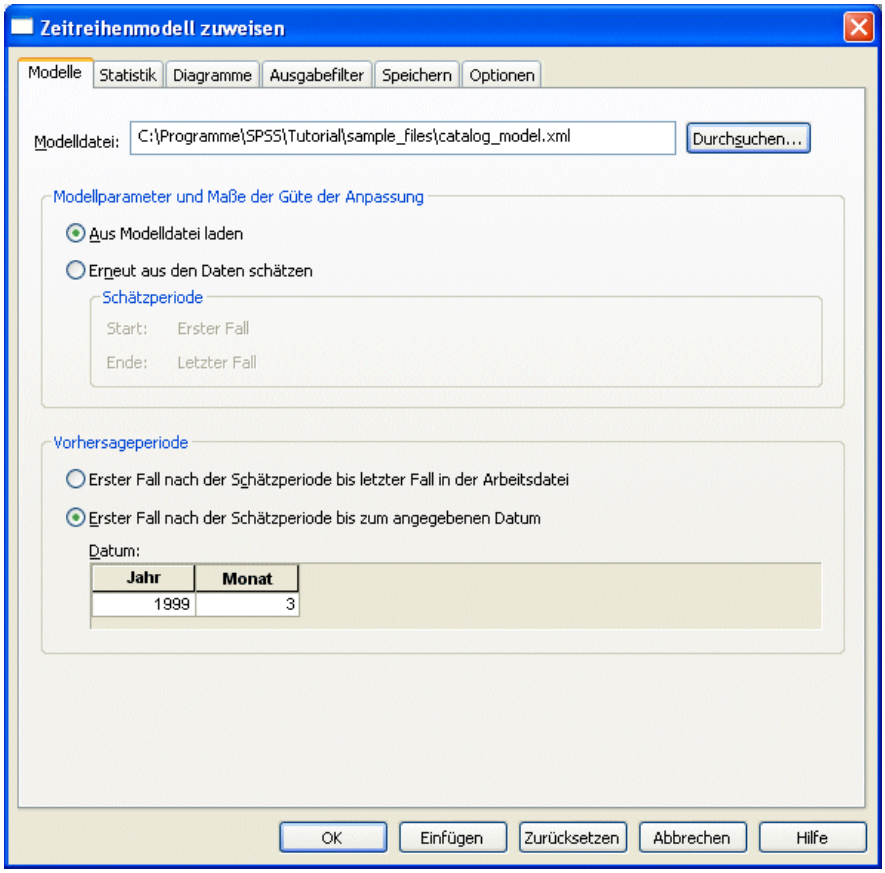
  Dadurch werden die Werte für *mail*, also die Anzahl der versendeten Kataloge, für jeden der drei Monate in der Vorhersageperiode um 2000 erhöht. Sie haben nun die Daten für den Test des ersten Szenarios vorbereitet und können die Analyse durchführen.

Durchführung der Analyse

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Zeitreihen
  Modelle zuweisen...

Abbildung 9-8
Dialogfeld "Zeitreihenmodell zuweisen"



- Klicken Sie auf Durchsuchen und wechseln Sie dann zu der Datei *catalog_model.xml* und wählen Sie sie aus (alternativ können Sie auch Ihre eigene Modelldatei aus dem vorhergehenden Beispiel auswählen). Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108.

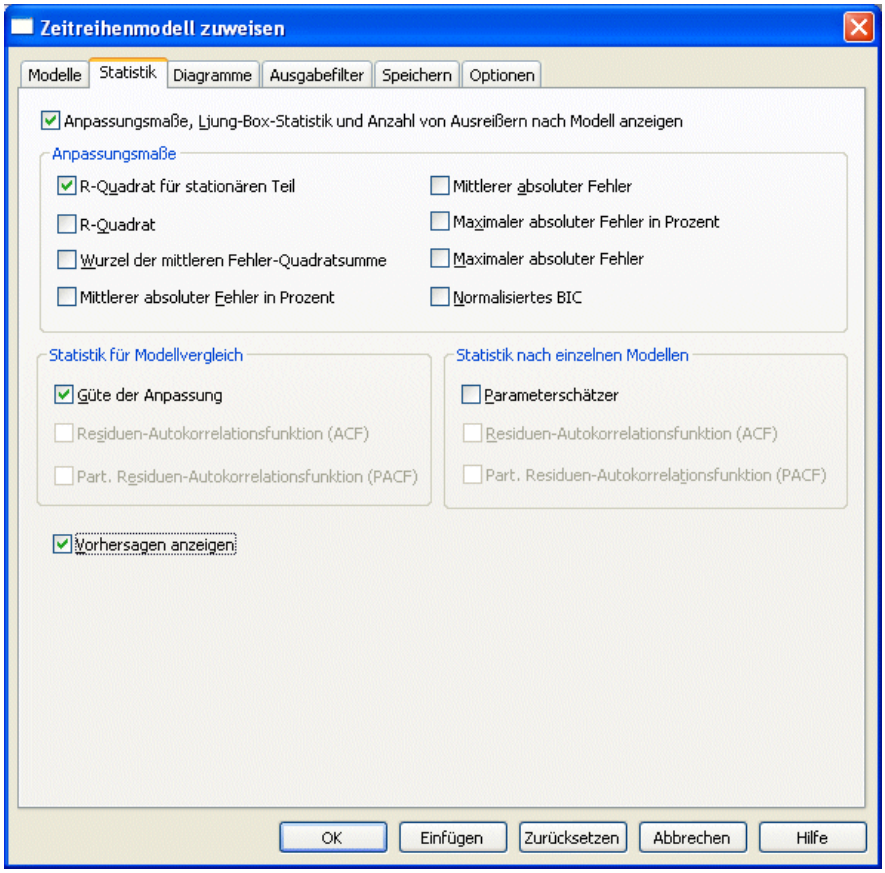
  Der Pfad von *catalog_model.xml* (oder der eigenen Modelldatei) wird nun auf der Registerkarte "Modelle" angezeigt.

- Wählen Sie im Gruppenfeld "Vorhersageperiode" die Option Erster Fall nach der Schätzperiode bis zum angegebenen Datum aus.

- Geben Sie im Datumsgitter für das Jahr 1999 und für den Monat 3 ein.

- Klicken Sie auf die Registerkarte Statistik.

Abbildung 9-9
Zeitreihenmodell zuweisen, Registerkarte "Statistik"



▶ Wählen Sie Vorhersagen anzeigen aus.

Hierdurch wird eine Tabelle von vorhergesagten Werten für die abhängige Variable ausgegeben.

▶ Klicken Sie im Dialogfeld "Zeitreihenmodell zuweisen" auf OK.

Abbildung 9-10
Vorhersagetabelle

Modell		JAN 1999	FEB 1999	MÄR 1999
Sales of Men's Clothing-Model_1	Vorhersagen	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

Die Tabelle der Vorhersagen enthält die vorhergesagten Werte der abhängigen Datenreihe unter Berücksichtigung der beiden Einflussvariablen *mail* und *phone* in der Vorhersageperiode. Die Tabelle enthält außerdem die obere Konfidenzgrenze (UCL) und die untere Konfidenzgrenze (LCL) für die Vorhersagen.

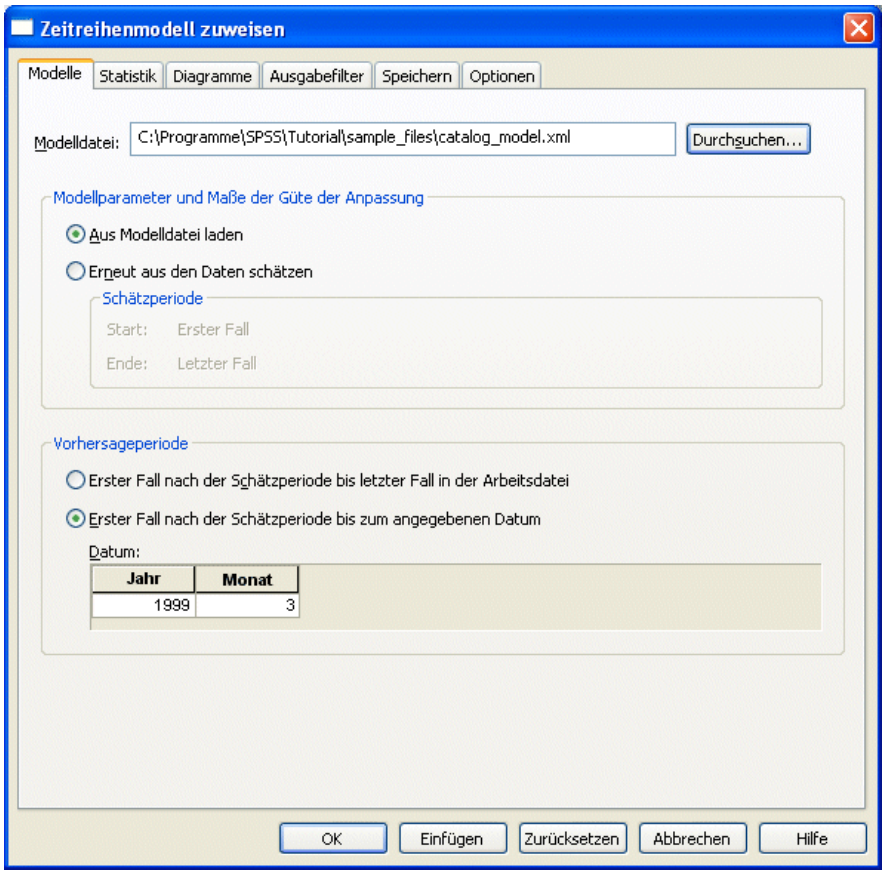
Sie haben die Umsatzvorhersage für das Szenario erstellt, bei dem Sie 2000 zusätzliche Kataloge pro Monat versenden. Nun bereiten Sie die Daten für das Szenario vor, bei dem Sie zusätzliche Telefonleitungen bereitstellen. Dazu müssen Sie die Variable *mail* auf die ursprünglichen Werte zurücksetzen und die Variable *phone* um 5 erhöhen. Sie können *mail*

zurücksetzen, indem Sie die Werte von *Vorhersagewert_mail_Modell_1* in der Vorhersageperiode kopieren und die aktuellen Werte von *mail* in der Vorhersageperiode mit diesen Werten überschreiben. Sie können die Anzahl der Telefonleitungen entweder direkt im Daten-Editor oder wie bei der Anzahl der Kataloge mithilfe des Dialogfelds “Variable berechnen” erhöhen. Der Wert muss für jeden Monat in der Vorhersageperiode um 5 erhöht werden.

Um die Analyse durchzuführen, öffnen Sie das Dialogfeld “Zeitreihenmodell zuweisen” wie folgt erneut:

- Klicken Sie in der Symbolleiste auf die Schaltfläche “Zuletzt verwendete Dialogfelder”.
- Wählen Sie Zeitreihenmodell zuweisen aus.

Abbildung 9-11
Dialogfeld “Zeitreihenmodell zuweisen”



- Klicken Sie im Dialogfeld “Zeitreihenmodell zuweisen” auf OK.

Abbildung 9-12
Tabellen der Vorhersagen für die beiden Szenarien

Forecast with more catalogs

Modell		JAN 1999	FEB 1999	MAR 1999
Sales of Men's Clothing-Model_1	Vorhersagen	25279.91	22064.72	21580.96
	UCL	27591.62	24376.42	23892.66
	LCL	22968.21	19753.02	19269.25

Forecast with more phone lines

Modell		JAN 1999	FEB 1999	MAR 1999
Sales of Men's Clothing-Model_1	Vorhersagen	23757.25	20542.06	20058.29
	UCL	26068.95	22853.76	22370.00
	LCL	21445.55	18230.35	17746.59

In den Tabellen der Vorhersagen für die beiden Szenarien zeigt sich, dass bei einer gesteigerten Anzahl versendeter Kataloge für jeden der drei vorhergesagten Monate ca. 1500 Dollar mehr Umsatz zu erwarten ist als bei mehr bereitgestellten Telefonleitungen. Auf der Grundlage dieser Analyse sollten die Ressourcen zum Versenden 2000 zusätzlicher Kataloge eingesetzt werden.

Kapitel
10

Saisonale Zerlegung

Entfernen der Saisonalität aus Verkaufsdaten

Eine Katalogfirma interessiert sich für die Modellierung des Aufwärtstrends im Absatz Ihres Herrenbekleidungssortiments anhand einer Menge von Einflussvariablen (wie der Anzahl der versendeten Kataloge und der Anzahl der Telefonleitungen, über die eine Bestellung möglich ist). Dazu hat das Unternehmen Daten über die monatlichen Verkaufszahlen für Herrenbekleidung über den Zeitraum von 10 Jahren gesammelt. Diese Informationen finden Sie in der Datei *catalog.sav*. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108.

Zur Durchführung einer Trendanalyse müssen alle möglicherweise in den Daten vorhandenen saisonalen Variationen entfernt werden. Dies lässt sich mit der Prozedur “Saisonale Zerlegung” problemlos erreichen.

Bestimmen und Festlegen der Periodizität

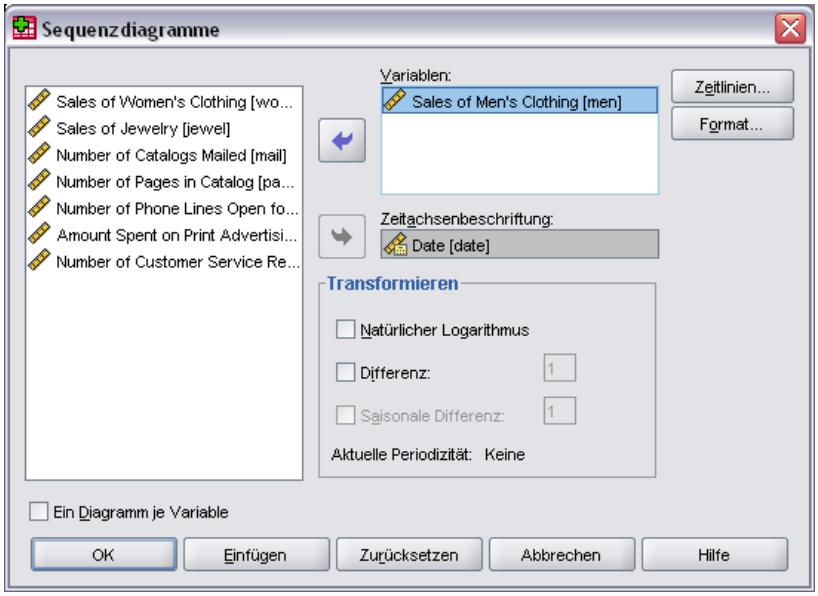
Für die Prozedur “Saisonale Zerlegung” muss die Arbeitsdatei eine periodische Datumskomponente enthalten, beispielsweise eine jährliche Periodizität von 12 (Monaten), eine wöchentliche Periodizität von 7 (Tagen) usw. Sie sollten zunächst die Zeitreihe grafisch darstellen, da die Untersuchung eines Zeitreihendiagramms oft zu einer brauchbaren Vermutung über die zugrunde liegende Periodizität führt.

So erhalten Sie ein Diagramm für den Absatz an Herrenbekleidung im Verlauf der Zeit:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

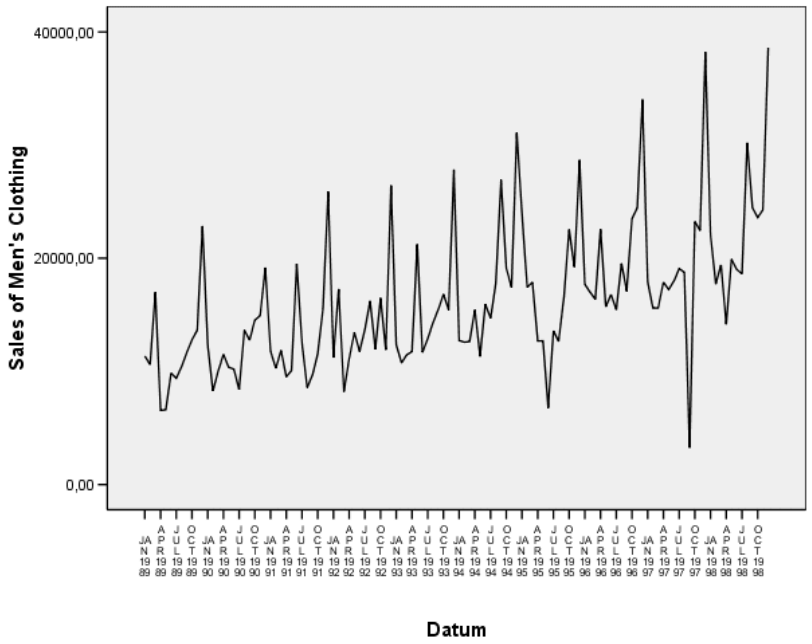
Analysieren
  Zeitreihen
    Sequenzdiagramme...

Abbildung 10-1
Dialogfeld "Sequenzdiagramme"



- Wählen Sie die Variable *Sales of Men's Clothing* aus und verschieben Sie sie in die Liste "Variablen".
- Wählen Sie die Variable *Date* und verschieben Sie sie in die Liste "Zeitachsenbeschriftung".
- Klicken Sie auf OK.

Abbildung 10-2
Absatz an Herrenbekleidung (in US-Dollar)



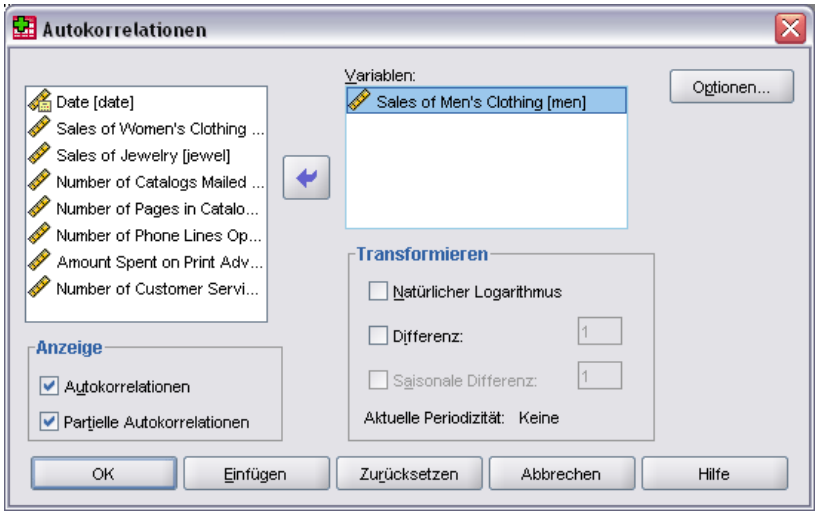
Die Zeitreihe weist eine Reihe von Spitzen auf, die sich jedoch nicht im gleichen Abstand zueinander zu befinden scheinen. Diese Ausgabe deutet darauf hin, dass die Zeitreihe eine periodische Komponente aufweist. Außerdem weist sie Fluktuationen auf, die nicht periodisch sind, der typische Fall bei Zeitreihen mit Echtzeit. Neben den kleineren Fluktuationen scheinen auch die signifikanten Spitzen weiter auseinanderzuliegen als nur ein paar Monate. In Anbetracht der saisonalen Natur des Absatzes, der normalerweise in der Weihnachtszeit Spitzenwerte aufweist, lässt sich schließen, dass die Zeitreihe wahrscheinlich eine jährliche Periodizität aufweist. Beachten Sie außerdem, dass die saisonalen Variationen im Einklang mit dem Aufwärtstrend der Zeitreihe zunehmen, was darauf hindeutet, dass die saisonalen Variationen proportional zum Niveau der Zeitreihe sein könnten. Dies wiederum legt nahe, dass es sich um ein multiplikatives und nicht um ein additives Modell handelt.

Die Untersuchung der Autokorrelationen und partiellen Autokorrelationen einer Zeitreihe erlaubt eine quantitativere Schlussfolgerung über die zugrunde liegende Periodizität.

▶ Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Autokorrelationen...

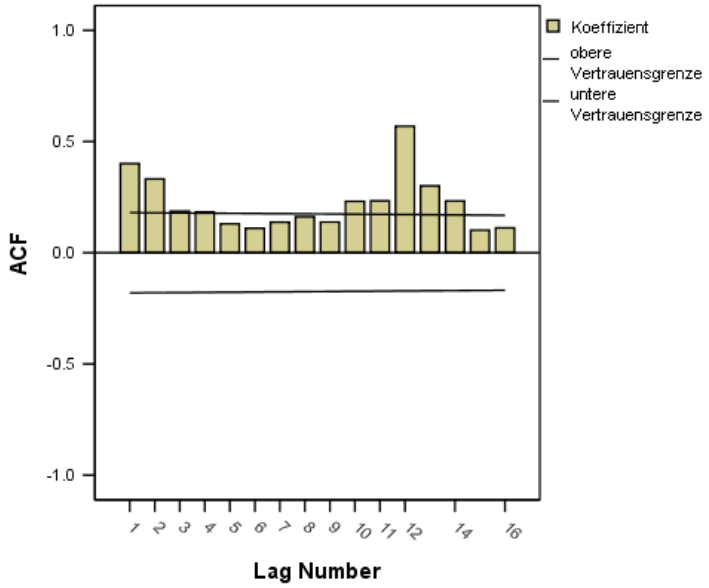
Abbildung 10-3
Dialogfeld "Autokorrelationen"



▶ Wählen Sie die Variable *Sales of Men's Clothing* aus und verschieben Sie sie in die Liste "Variablen".

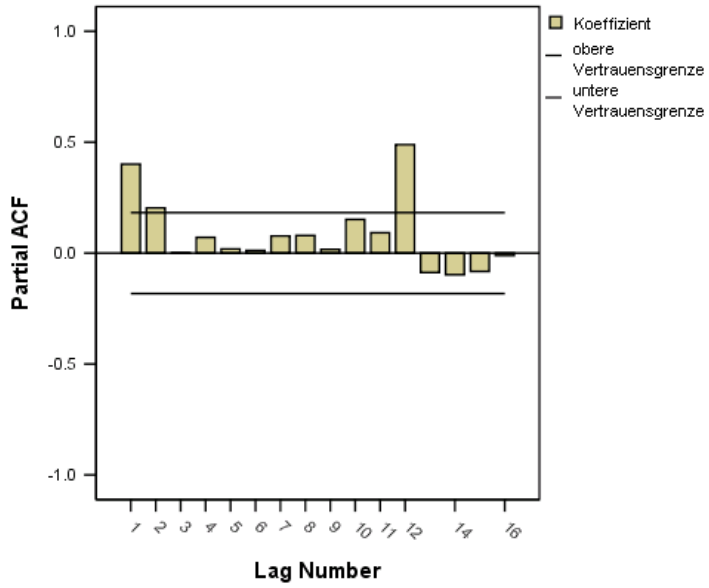
▶ Klicken Sie auf OK.

Abbildung 10-4
Autokorrelationsdiagramm für Herrenbekleidung



Die Autokorrelationsfunktion zeigt eine signifikante Spitze bei Lag 1 mit einem langen exponentiellen Schwanz: ein typisches Muster für Zeitreihen. Die signifikante Spitze bei Lag 12 deutet auf das Vorhandensein einer jährlichen saisonalen Komponente in den Daten hin. Die Untersuchung der partiellen Autokorrelationsfunktion erlaubt eine eindeutigere Schlussfolgerung.

Abbildung 10-5
Partielles Autokorrelationsdiagramm für Herrenbekleidung



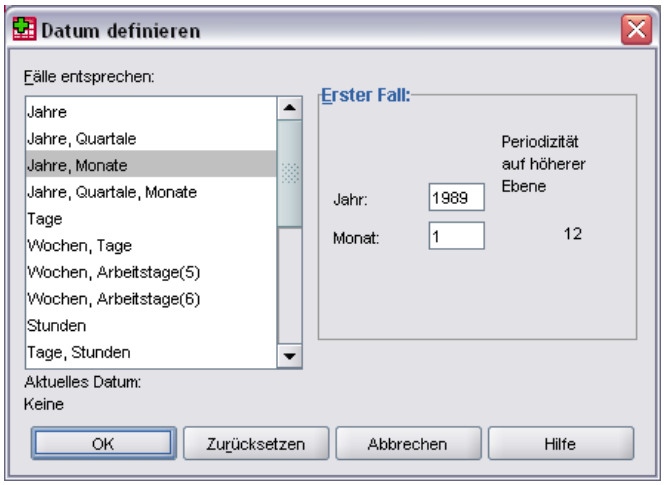
Die signifikante Spitze bei Lag 12 in der der partiellen Autokorrelationsfunktion bestätigt das Vorhandensein einer jährlichen saisonalen Komponente in den Daten.

So legen Sie eine jährliche Periodizität fest:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Daten
  Datum definieren...

Abbildung 10-6
Dialogfeld "Datum definieren"



- Wählen Sie Jahre, Monate aus der Liste "Fälle entsprechen:" aus.
- Geben Sie 1989 als Jahr und 1 als Monat ein.
- Klicken Sie auf OK.

  Dadurch wird die Periodizität auf 12 gesetzt und eine Menge von Datumsvariablen für die Arbeit mit den Prozeduren von Trends erstellt.

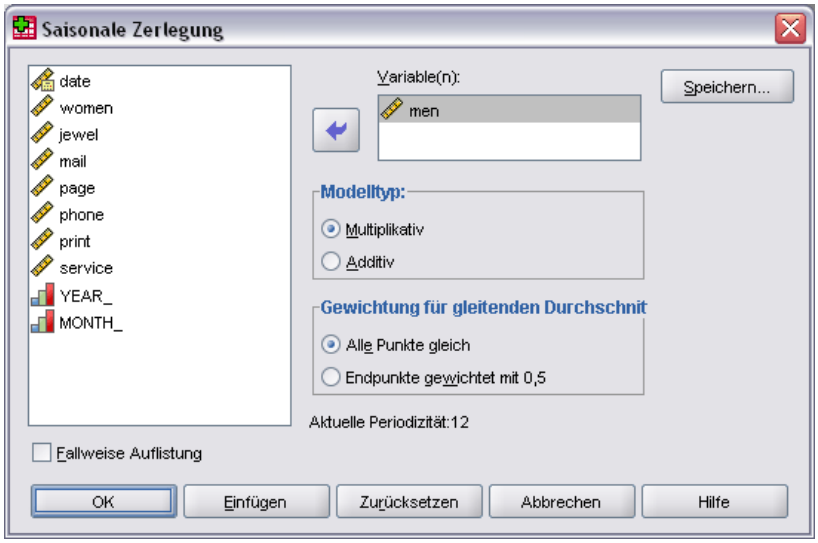
Durchführung der Analyse

So führen Sie die Prozedur "Saisonale Zerlegung" durch:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

  Analysieren
  Zeitreihen
  Saisonale Zerlegung...

Abbildung 10-7
Dialogfeld "Saisonale Zerlegung"



- Klicken Sie mit der rechten Maustaste auf eine beliebige Stelle in der Quellvariablenliste und wählen Sie im Kontextmenü die Option Variablennamen anzeigen aus.
- Wählen Sie die Variable *men* aus und verschieben Sie sie in die Liste "Variablen".
- Wählen Sie im Gruppenfeld "Modelltyp" die Option Multiplikativ.
- Klicken Sie auf OK.

Interpretation der Ausgabe

Die Prozedur "Saisonale Zerlegung" erstellt vier neue Variablen für jede von der Prozedur analysierte ursprüngliche Variable. In der Standardeinstellung werden die neuen Variablen zur Arbeitsdatei hinzugefügt. Die Namen der neuen Zeitreihen beginnen mit folgenden Präfixen:

SAF. Faktoren für die Saisonbereinigung, die für die saisonale Variation stehen. Beim multiplikativen Modell steht der Wert 1 für die Abwesenheit von saisonaler Variation; beim additiven Modell steht der Wert 0 für die Abwesenheit von saisonaler Variation.

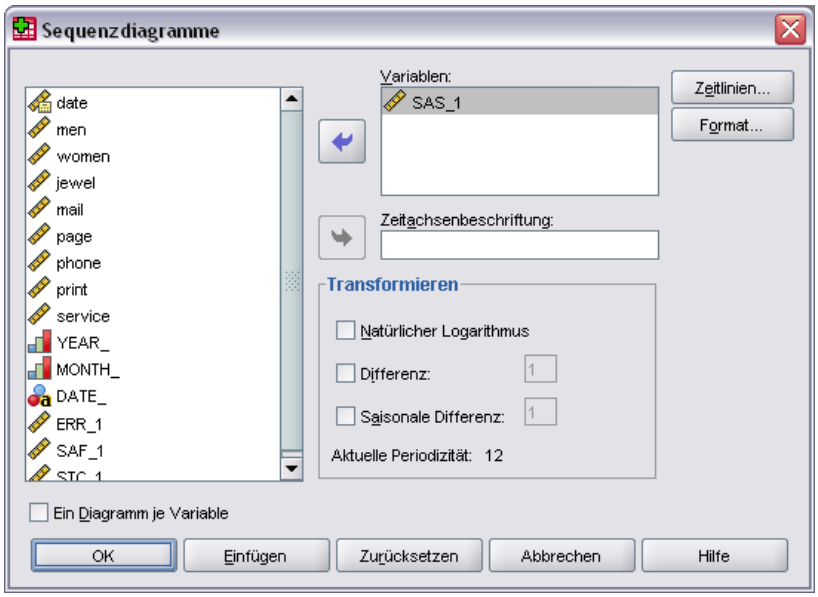
SAS. Saisonbereinigte Zeitreihe, die für die ursprüngliche Zeitreihe nach der Entfernung der saisonalen Variationen steht. Wenn mit einer saisonbereinigten Zeitreihe gearbeitet wird, kann beispielsweise eine Trend-Komponente isoliert und unabhängig von etwaigen saisonalen Komponenten analysiert werden.

STC. Geglättete Trend-Zyklus-Komponente, eine geglättete Version der saisonbereinigten Zeitreihe, die sowohl Trend-Komponenten als auch zyklische Komponenten aufweist.

ERR. Die Residuum-Komponente der Zeitreihe für eine bestimmte Beobachtung.

Beim vorliegenden Fall ist die saisonbereinigte Zeitreihe am besten geeignet, da sie für die ursprüngliche Zeitreihe nach Entfernung der saisonalen Variationen steht.

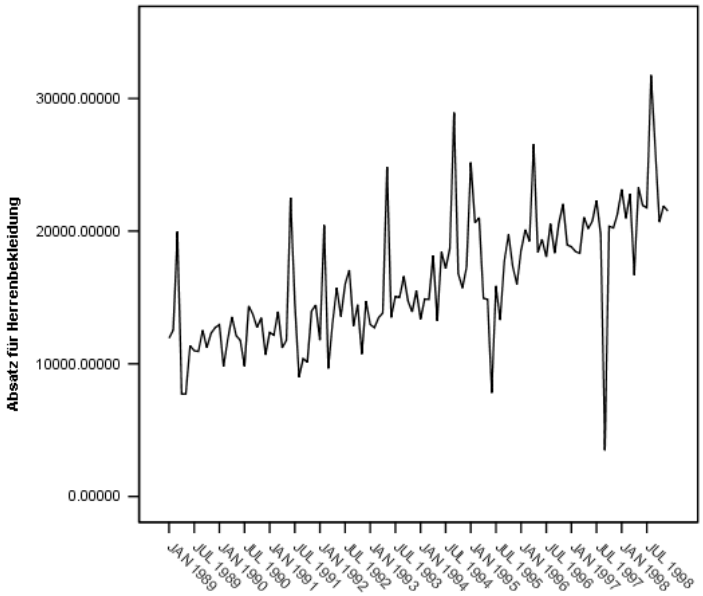
Abbildung 10-8
Dialogfeld "Sequenzdiagramme"



So stellen Sie die saisonbereinigte Zeitreihe grafisch dar :

- Öffnen Sie das Dialogfeld "Sequenzdiagramme".
- Klicken Sie auf Zurücksetzen, um frühere Einstellungen zu löschen.
- Klicken Sie mit der rechten Maustaste auf eine beliebige Stelle in der Quellvariablenliste und wählen Sie im Kontextmenü die Option Variablennamen anzeigen aus.
- Wählen Sie die Variable *SAS_1* aus und verschieben Sie sie in die Liste "Variablen".
- Klicken Sie auf OK.

Abbildung 10-9
Saisonbereinigte Zeitreihe



Die saisonbereinigte Zeitreihe zeigt einen klaren Aufwärtstrend. Es ist eine Reihe von Spitzen vorhanden; sie erscheinen jedoch in zufälligen Abständen ohne Anzeichen für ein jährliches Muster.

Zusammenfassung

Mit der Prozedur “Saisonale Zerlegung” wurde die saisonale Komponente einer periodischen Zeitreihe entfernt, um eine Zeitreihe zu erstellen, die sich besser für die Trendanalyse eignet. Die Untersuchung der Autokorrelationen und partiellen Autokorrelationen war hilfreich für die Ermittlung der zugrunde liegenden Periodizität (in diesem Fall: jährlich).

Verwandte Prozeduren

Die Prozedur “Saisonale Zerlegung” dient zur Entfernung einer einzelnen saisonalen Komponente aus einer periodischen Zeitreihe.

- Um eine tiefer gehende Analyse der Periodizität einer Zeitreihe durchzuführen, als es mit der partiellen Korrelationsfunktion möglich ist, sollten Sie die Prozedur “Spektraldiagramme” verwenden. Weitere Informationen finden Sie unter Kapitel 11.

Kapitel
11

Spektraldiagramme

Verwendung von Spektraldiagrammen zur Überprüfung der Erwartungen hinsichtlich der Periodizität

Zeitreihen, die für Absätze im Einzelhandel stehen, weisen normalerweise eine zugrunde liegende jährliche Periodizität auf, die auf den üblichen Anstieg des Absatzes in der Vorweihnachtszeit zurückzuführen ist. Wenn Absatzprojektionen erstellt werden sollen, muss ein Modell der Zeitreihe konstruiert werden, was wiederum bedeutet, dass etwaige periodische Komponenten identifiziert werden müssen. Ein Diagramm der Zeitreihe offenbart möglicherweise nicht immer die jährliche Periodizität, da Zeitreihen Zufallsfluktuationen aufweisen, die häufig die zugrunde liegende Struktur maskieren.

Die monatlichen Absatzdaten für ein Versandhaus sind in der Datei *catalog.sav* gespeichert. Für weitere Informationen siehe Beispieldateien in Anhang D auf S. 108. Bevor Sie mit den Absatzprojektionen fortfahren, möchten Sie bestätigt wissen, dass die Absatzdaten eine jährliche Periodizität aufweisen. Ein Diagramm der Zeitreihe zeigt viele Spitzen mit unregelmäßigem Abstand, sodass eine etwaige zugrunde liegende Periodizität nicht ersichtlich ist. Verwenden Sie die Prozedur “Spektraldiagramme”, um eine etwaige Periodizität in den Absatzdaten zu identifizieren.

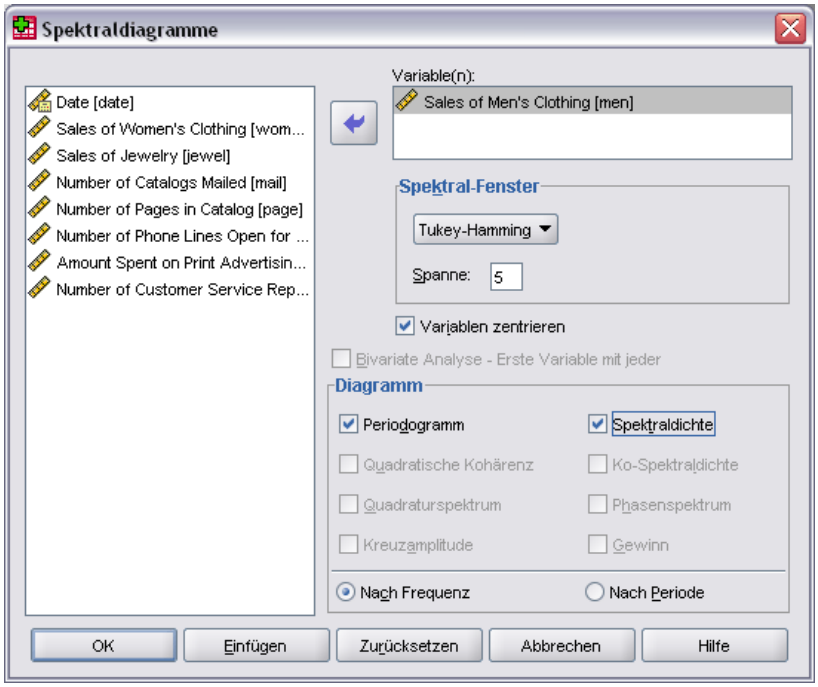
Durchführung der Analyse

So führen Sie die Prozedur “Spektraldiagramme” durch:

- Wählen Sie die folgenden Befehle aus den Menüs aus:

Analysieren
  Zeitreihen
    Spektralanalyse...

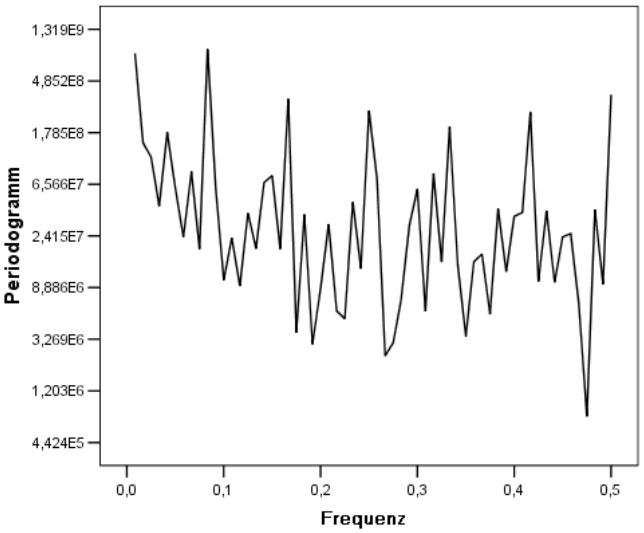
Abbildung 11-1
Dialogfeld "Spektraldiagramme"



- Wählen Sie die Variable *Sales of Men's Clothing* aus und verschieben Sie sie in die Liste "Variablen".
- Wählen Sie im Gruppenfeld "Diagramm" die Option Spektraldichte aus.
- Klicken Sie auf OK.

Zum Verständnis des Periodogramms und der Spektraldichte

Abbildung 11-2
Periodogramm



Das Periodogramm zeigt eine Folge von Spitzen, die aus dem Hintergrundrauschen hervortreten. Die Spitze mit der niedrigsten Frequenz liegt bei einer Frequenz von knapp unter 0,1. Sie vermuten, dass die Daten eine jährliche periodische Komponente aufweisen. Daher müssen Sie den Beitrag berücksichtigen, den eine jährliche Komponente zu dem Periodogramm leisten würde. Jeder Datenpunkt in der Zeitreihe steht für einen Monat. Daher entspricht eine jährliche Periodizität im aktuellen Daten-Set einer Periode von 12. Da Periode und Frequenz Kehrwerte voneinander sind, entspricht eine Periode von 12 einer Frequenz von 1/12 (bzw. 0,083). Eine jährliche Komponente impliziert also eine Spitze im Periodogramm bei 0,083, was mit dem Vorhandensein der Spitze knapp unterhalb einer Frequenz von 0,1 konsistent zu sein scheint.

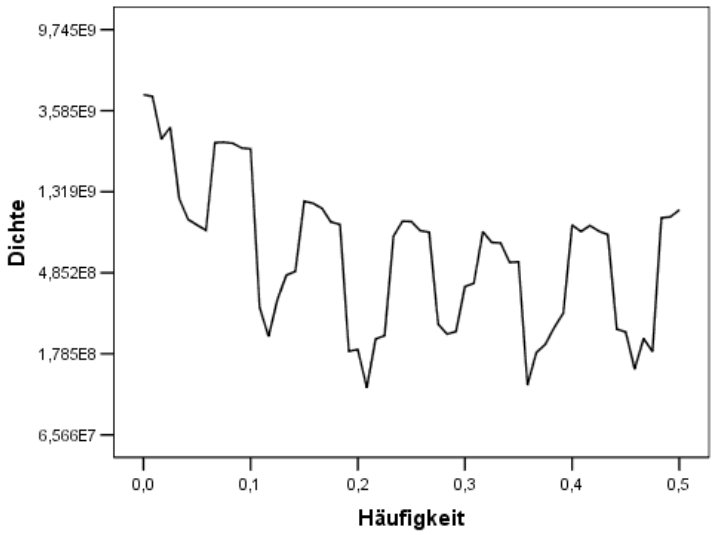
Abbildung 11-3
Tabelle für univariate Statistiken

Name der Reihe: men

	Häufigkeit	Sinus-Transformation	Kosinus-Transformation	Periodogramm	Spektral dichte schätzung
1	,00000	,000	16242,813	,000	4,361E+09
2	,00833	-3696,643	370,153	828131182	4,278E+09
3	,01667	-1496,277	454,831	146743038	2,526E+09
4	,02500	-1336,400	252,087	110970821	2,921E+09
5	,03333	-662,146	529,734	43143315,6	1,210E+09
6	,04167	-1654,614	501,571	179359225	935924937
7	,05000	-784,814	-636,729	61281367,3	875492029
8	,05833	-335,646	532,062	23744855,0	820375352
9	,06667	-1094,178	-451,489	84064108,9	2,420E+09
10	,07500	264,554	492,876	18774933,5	2,429E+09
11	,08333	-3053,934	2370,483	896742149	2,401E+09
12	,09167	-978,882	-287,035	62435897,6	2,263E+09
13	,10000	-403,128	93,036	10270064,4	2,235E+09

Die Tabelle für univariate Statistiken enthält die Datenpunkte, die für die grafische Darstellung des Periodogramms verwendet wurden. Beachten Sie: Für Frequenzen von weniger als 0,1 erscheint der größte Wert in der Spalte *Periodogramm* bei einer Frequenz von 0,08333 – das ist genau der Wert, der zu erwarten ist, wenn eine jährliche periodische Komponente vorliegt. Diese Daten bestätigen die Identifikation der Spitze mit der niedrigsten Frequenz mit einer jährlichen periodischen Komponente. Doch was ist mit den anderen Spitzen bei höheren Frequenzen?

Abbildung 11-4
Spektraldichte



Die restlichen Spitzen werden am besten mit der Spektraldichtefunktion analysiert, bei der es sich einfach um eine geglättete Version des Periodogramms handelt. Die Glättung bietet ein Mittel zur Beseitigung des Hintergrundrauschens aus einem Periodogramm, wodurch die zugrunde liegende Struktur besser isoliert werden kann.

Die Spektraldichte besteht aus fünf unterschiedlichen Spitzen, die in gleichen Abständen erscheinen. Die Spitze mit der niedrigsten Frequenz entspricht einfach der geglätteten Version der Spitze bei 0,08333. Um die Bedeutung der vier Spitzen bei höheren Frequenzen zu verstehen, müssen Sie bedenken, dass das Periodogramm berechnet wurde, indem die Zeitreihen als Summe aus Kosinus- und Sinusfunktion modelliert wurde. Periodische Komponenten, die die Gestalt einer Sinus- oder Kosinusfunktion haben (sinusförmig) erscheinen im Periodogramm als einzelne Spitzen. Periodische Komponenten, die nicht sinusförmig sind, erscheinen als Reihen von Spitzen mit verschiedenen Höhen im gleichen Abstand, wobei die Spitze mit der niedrigsten Frequenz in der Reihe bei der Frequenz der periodischen Komponente auftritt. Daher weisen die vier Spitzen bei höheren Frequenzen in der Spektraldichte lediglich darauf hin, dass die jährliche periodische Komponente nicht sinusförmig ist.

Sie haben nun alle erkennbaren Strukturen im Spektraldichtediagramm erklärt und schließen daraus, dass die Daten eine einzelne periodische Komponente mit einer Periode von 12 Monaten enthalten.

Zusammenfassung

Mit der Prozedur “Spektraldiagramme” haben Sie die Existenz einer jährlichen periodischen Komponente einer Zeitreihe bestätigt und verifiziert, dass keine anderen signifikanten Periodizitäten vorliegen. Die Spektraldichte hat sich bei der Offenlegung der zugrunde liegenden Struktur als nützlicher als das Periodogramm erwiesen, da die Spektraldichte die von der nichtperiodischen Komponente der Daten verursachten Fluktuationen glättet.

Verwandte Prozeduren

Die Prozedur “Spektraldiagramme” kann zur Identifizierung der periodischen Komponenten einer Zeitreihen verwendet werden.

- Um eine periodische Komponente aus einer Zeitreihe zu entfernen – beispielsweise um eine Trendanalyse durchzuführen –, können Sie die Prozedur “Saisonale Zerlegung” verwenden. Unter Kapitel 10 finden Sie weitere Einzelheiten.

Anhang

A

Maße für die Anpassungsgüte

In diesem Abschnitt finden Sie Definitionen der Maße für die Anpassungsgüte, die bei der Zeitreihenmodellierung verwendet werden.

- **R-Quadrat für stationären Teil.** Ein Maß, das den stationären Teil des Modells mit einem einfachen Mittelwert-Modell vergleicht. Dieses Maß ist dem gewöhnlichen R-Quadrat vorzuziehen, wenn ein Trend oder ein saisonales Muster vorliegt. R-Quadrat für den stationären Teil kann auch negativ sein, es nimmt Werte zwischen minus unendlich und 1 an. Negative Werte bedeuten, dass das betrachtete Modell schlechter ist als das Basismodell. Positive Werte bedeuten, dass das betrachtete Modell besser ist als das Basismodell.
- **R-Quadrat.** Ein Schätzer für den Anteil der Gesamtvariation der Zeitreihe, der durch das Modell erklärt wird. Dieses Maß ist sehr nützlich, wenn die Zeitreihe stationär ist. R-Quadrat kann auch negativ sein, es nimmt Werte zwischen minus unendlich und 1 an. Negative Werte bedeuten, dass das betrachtete Modell schlechter ist als das Basismodell. Positive Werte bedeuten, dass das betrachtete Modell besser ist als das Basismodell.
- **RMSE.** Steht für Root Mean Square Error, die Wurzel des mittleren quadratischen Fehlers. Die Quadratwurzel des mittleren Fehlerquadrats. Ein Maß dafür, wie stark eine abhängige Zeitreihe von ihrem durch das Modell vorhergesagten Niveau abweicht, und zwar ausgedrückt in derselben Maßeinheit wie die abhängige Zeitreihe.
- **MAPE.** Mittlerer absoluter Fehler in Prozent. Ein Maß dafür, wie stark eine abhängige Zeitreihe von ihrem durch das Modell vorhergesagten Niveau abweicht. Es ist unabhängig von den verwendeten Maßeinheiten und kann daher verwendet werden, um Zeitreihen mit unterschiedlichen Einheiten zu vergleichen.
- **MAE.** Mean Absolute Error, also mittlerer absoluter Fehler bzw. mittlerer Betrag des Fehlers. Er misst, wie stark die Zeitreihe von ihrem durch das Modell vorhergesagten Niveau abweicht. MAE wird in derselben Maßeinheit angegeben wie die ursprüngliche Zeitreihe.
- **MaxAPE.** Maximaler absoluter Fehler in Prozent (Maximum Absolute Percentage Error, also maximaler Betrag des relativen Fehlers). Dies ist der größte vorhergesagte Fehler, ausgedrückt in Prozent. Dieses Maß hilft dabei, sich ein Worst-Case-Szenario für die Vorhersagen vorzustellen.
- **MaxAE.** Maximaler absoluter Fehler (Maximum Absolute Error, also maximaler Betrag des Fehlers). Dies ist der größte vorhergesagte Fehler, ausgedrückt in derselben Maßeinheit wie die abhängige Zeitreihe. Genau wie MaxAPE hilft er dabei, sich ein Worst-Case-Szenario für die Vorhersagen vorzustellen. Der maximale absolute Fehler und der maximale absolute Fehler in Prozent können an verschiedenen Punkten in der Zeitreihe auftreten, beispielsweise wenn der absolute Fehler für einen großen Zeitreihenwert geringfügig größer ist als der absolute Fehler für einen kleinen Zeitreihenwert. In diesem Fall tritt der maximale absolute

Fehler beim größeren Zeitreihenwert und der maximale absolute Fehler in Prozent beim kleineren Zeitreihenwert auf.

- **Normalisiertes BIC.** Normalisiertes Bayes-Informationskriterium (BIC). Ein allgemeines Maß der insgesamt erreichten Güte der Anpassung, das auch die Komplexität des Modells zu berücksichtigen versucht. Es ist ein Wert, der auf dem mittleren quadratischen Fehler beruht und eine Penalisierung für die Anzahl der Modellparameter und die Länge der Zeitreihe enthält. Die Penalisierung neutralisiert die Überlegenheit von Modellen mit einer größeren Anzahl von Parametern und macht die Statistik damit gut vergleichbar für verschiedene Modelle derselben Zeitreihe.

Anhang

B

Ausreißertypen

In diesem Abschnitt finden Sie Definitionen der Ausreißertypen, die bei der Zeitreihenmodellierung verwendet werden.

- **Additiv.** Ein Ausreißer, der eine einzelne Beobachtung beeinflusst. So könnte z. B. ein Fehler bei der Datenkodierung als additiver Ausreißer identifiziert werden.
- **Niveauverschiebung.** Ein Ausreißer, der beginnend an einer bestimmten Stelle der Zeitreihe alle Beobachtungen um eine Konstante verschiebt. Eine Niveauverschiebung könnte sich aus einem Strategiewechsel ergeben.
- **Innovativ (Neuerung).** Ein Ausreißer, der als Zuwachs zum Rauschen an einer bestimmten Stelle der Zeitreihe wirkt. Bei stationären Zeitreihen beeinflusst ein innovativer Ausreißer mehrere Beobachtungen. Bei nichtstationären Zeitreihen kann er alle Beobachtungen ab einer bestimmten Stelle der Zeitreihe beeinflussen.
- **Transient.** Ein Ausreißer, dessen Auswirkung exponentiell gegen null strebt.
- **Saisonal additiv.** Ein Ausreißer, der eine bestimmte Beobachtung beeinflusst sowie alle folgenden Beobachtungen, die von dieser eine oder mehrere Perioden entfernt sind. Alle diese Beobachtungen werden in gleicher Weise beeinflusst. Ein saisonal additiver Ausreißer könnte auftreten, wenn ab einem bestimmten Jahr die Umsätze jeden Januar höher sind.
- **Lokaler Trend.** Ein Ausreißer, der an einer bestimmten Stelle der Zeitreihe einen lokalen Trend auslöst.
- **Additiver Bereich.** Eine Gruppe von zwei oder mehr aufeinanderfolgenden additiven Ausreißern. Wenn dieser Typ von Ausreißern gewählt wird, werden neben Gruppen (Bereichen) von Ausreißern auch einzelne additive Ausreißer entdeckt.

Anhang

C

Übersicht über ACF-/PACF-Diagramme

Die hier gezeigten Diagramme gehören zu reinen oder theoretischen ARIMA-Prozessen. Hier einige allgemeine Richtlinien zur Identifizierung des Prozesses:

- Nichtstationäre Zeitreihen weisen eine ACF auf, die über mindestens ein halbes Dutzend Lags signifikant bleibt, anstatt sich schnell dem Wert 0 anzunähern. Sie müssen eine solche Zeitreihe vor dem Identifizieren des Prozesses differenzieren, bis sie stationär ist.
- Autoregressive Prozesse weisen eine exponentiell abnehmende ACF und Spikes im ersten Lag (oder den ersten Lags) der PACF auf. Die Anzahl der Spikes gibt die Ordnung der Autoregression an.
- Prozesse mit gleitendem Durchschnitt Verfahren weisen Spikes im ersten Lag (oder den ersten Lags) der ACF und eine exponentiell abnehmende PACF auf. Die Anzahl der Spikes gibt die Ordnung des gleitenden Durchschnitts an.
- Gemischte (ARMA-)Prozesse weisen normalerweise eine exponentielle Abnahme sowohl bei der ACF als auch bei der PACF auf.

In der Identifizierungsphase brauchen Sie sich nicht um das Vorzeichen der ACF bzw. PACF oder um die Geschwindigkeit zu kümmern, mit der eine exponentiell abnehmende ACF bzw. PACF sich dem Wert 0 annähert. Diese hängt vom Vorzeichen und dem tatsächlichen Wert der AR- und MA-Koeffizienten ab. In einigen Fällen wechselt eine exponentiell abnehmende ACF zwischen positiven und negativen Werten.

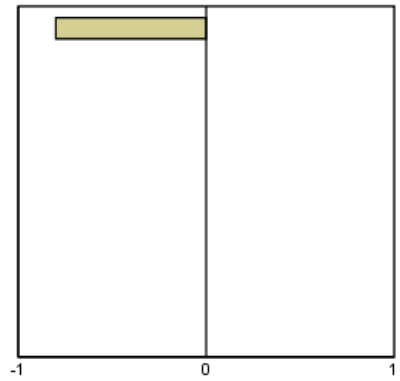
ACF- und PACF-Diagramme aus echten Daten sind nie so sauber wie die hier gezeigten Diagramme. Sie müssen lernen, aus dem jeweiligen Diagramm das Wesentliche herauszusuchen. Überprüfen Sie stets die ACF und PACF der Residuen, für den Fall, dass Ihre Identifizierung falsch ist. Beachten Sie:

- Saisonale Prozesse zeigen diese Muster in den saisonalen Lags (den Mehrfachen der Saison-Periode).
- Sie dürfen nichtsignifikante Werte als 0 behandeln, d. h., Sie können Werte ignorieren, die im Diagramm innerhalb der Konfidenzintervalle liegen. Sie müssen sie jedoch nicht ignorieren, insbesondere, wenn Sie das Muster der statistisch signifikanten Werte fortsetzen.
- Einzelne Autokorrelationen sind schon durch Zufall statistisch signifikant. Sie können eine statistisch signifikante Autokorrelation ignorieren, wenn sie isoliert vorkommt, insbesondere in hohen Lags, und wenn sie nicht in einem saisonalen Lag auftritt.

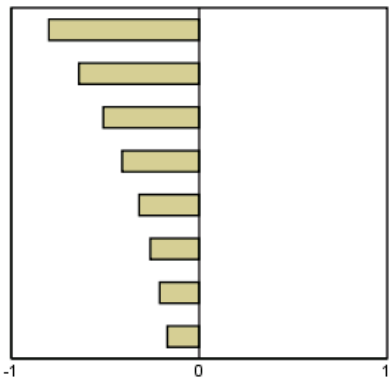
Detailliertere Erläuterungen zu ACF- und PACF-Diagrammen finden Sie in jedem beliebigen Text zur ARIMA-Analyse.

ARIMA(0,0,1), θ>0

ACF

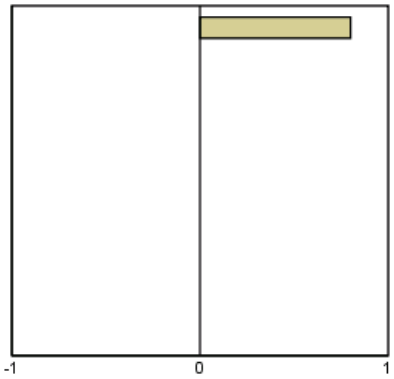


PACF

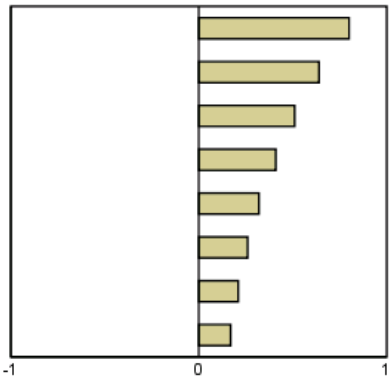


ARIMA(0,0,1), θ<0

ACF

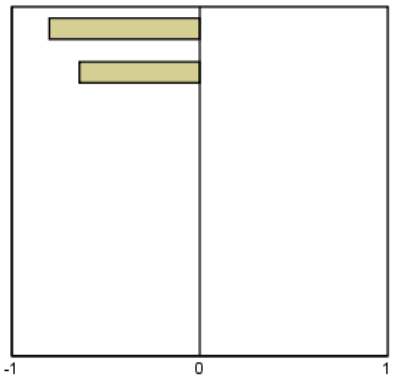


PACF

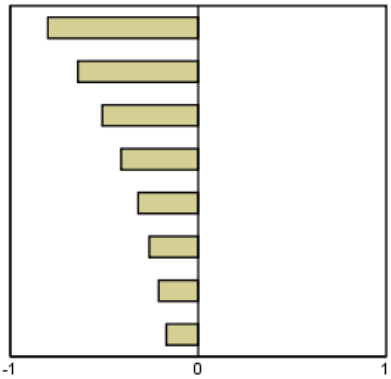


ARIMA(0,0,2), $\theta_1\theta_2>0$

ACF

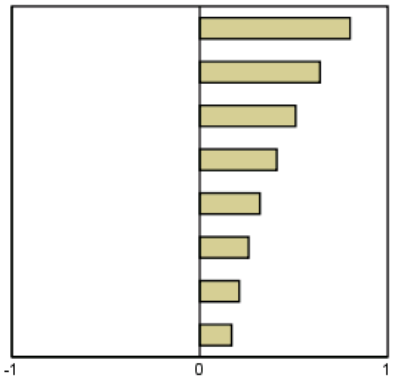


PACF

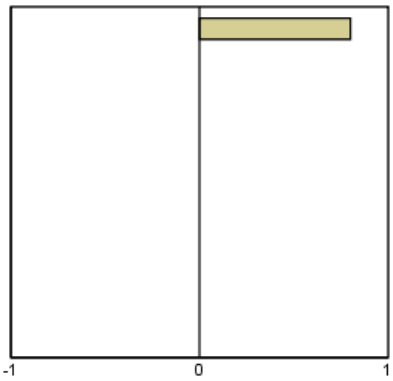


ARIMA(1,0,0), $\phi>0$

ACF

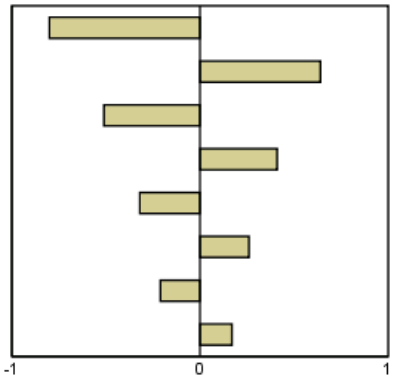


PACF

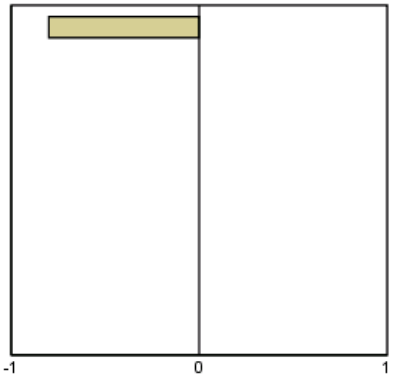


ARIMA(1,0,0), $\phi<0$

ACF

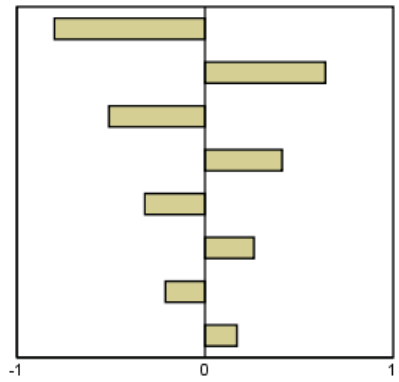


PACF

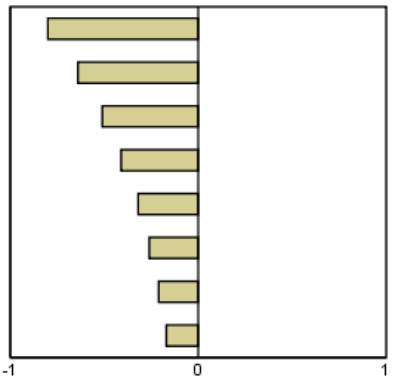


ARIMA(1,0,1), $\phi<0$, $\theta>0$

ACF

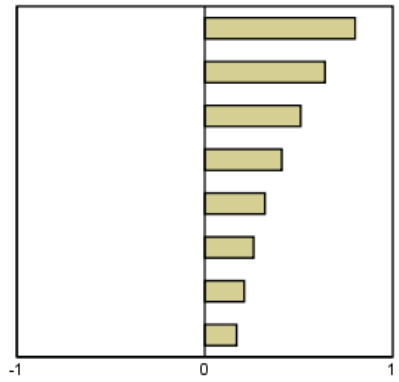


PACF

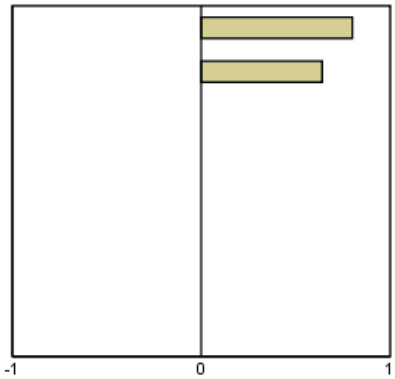


ARIMA(2,0,0), $\phi_1\phi_2>0$

ACF

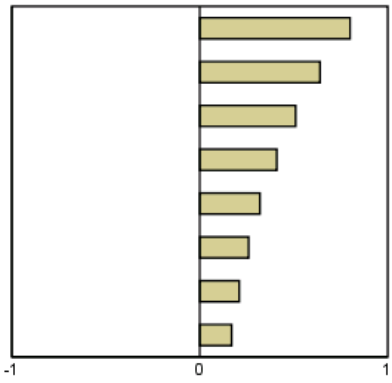


PACF



ARIMA(0,1,0) (integrierte Zeitreihe)

ACF



Anhang

D

Beispieldateien

Die zusammen mit dem Produkt installierten Beispieldateien finden Sie im Unterverzeichnis *Samples* des Installationsverzeichnisses.

Beschreibungen

Im Folgenden finden Sie Kurzbeschreibungen der in den verschiedenen Beispielen in der Dokumentation verwendeten Beispieldateien:

- **accidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die alters- und geschlechtsabhängige Risikofaktoren für Autounfälle in einer bestimmten Region untersucht. Jeder Fall entspricht einer Kreuzklassifikation von Alterskategorie und Geschlecht.
- **adl.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die Vorteile einer vorgeschlagenen Therapieform für Schlaganfallpatienten zu ermitteln. Ärzte teilten weibliche Schlaganfallpatienten nach dem Zufallsprinzip jeweils einer von zwei Gruppen zu. Die erste Gruppe erhielt die physische Standardtherapie, die zweite erhielt eine zusätzliche Emotionaltherapie. Drei Monate nach den Behandlungen wurden die Fähigkeiten der einzelnen Patienten, übliche Alltagsaktivitäten auszuführen, als ordinale Variablen bewertet.
- **advert.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Einzelhändlers geht, die Beziehungen zwischen den in Werbung investierten Beträgen und den daraus resultierenden Umsätzen zu untersuchen. Zu diesem Zweck hat er die Umsätze vergangener Jahre und die zugehörigen Werbeausgaben zusammengestellt.
- **aflatoxin.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests von Maisernten auf Aflatoxin geht, ein Gift, dessen Konzentration stark zwischen und innerhalb von Ernteerträgen schwankt. Ein Kornverarbeitungsbetrieb hat aus 8 Ernteerträgen je 16 Proben erhalten und das Aflatoxinniveau in Teilen pro Milliarde (parts per billion, PPB) gemessen.
- **aflatoxin20.sav.** Diese Datendatei enthält die Aflatoxinmessungen aus jeder der 16 Stichproben aus den Erträgen 4 und 8 der Datendatei *aflatoxin.sav*.
- **anorectic.sav.** Bei der Ausarbeitung einer standardisierten Symptomatologie anorektischen/bulimischen Verhaltens führten Forscher (Van der Ham, Meulman, Van Strien, als auch Van Engeland, 1997) eine Studie mit 55 Jugendlichen mit bekannten Ess-Störungen durch. Jeder Patient wurde vier Mal über einen Zeitraum von vier Jahren untersucht, es fanden also insgesamt 220 Beobachtungen statt. Bei jeder Beobachtung erhielten die Patienten Scores für jedes von 16 Symptomen. Die Symptomwerte fehlen für Patient 71

zum Zeitpunkt 2, Patient 76 zum Zeitpunkt 2 und Patient 47 zum Zeitpunkt 3, wodurch 217 gültige Beobachtungen verbleiben.

- **autoaccidents.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Versicherungsanalysten geht, ein Modell zur Anzahl der Autounfälle pro Fahrer unter Berücksichtigung von Alter und Geschlecht zu erstellen. Jeder Fall stellt einen Fahrer dar und erfasst das Geschlecht des Fahrers, sein Alter in Jahren und die Anzahl der Autounfälle in den letzten fünf Jahren.
- **band.sav.** Diese Datendatei enthält die hypothetischen wöchentlichen Verkaufszahlen von CDs für eine Musikgruppe. Daten für drei mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **bankloan.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Bank geht, den Anteil der nicht zurückgezahlten Kredite zu reduzieren. Die Datei enthält Informationen zum Finanzstatus und demografischen Hintergrund von 850 früheren und potenziellen Kunden. Bei den ersten 700 Fällen handelt es sich um Kunden, denen bereits ein Kredit gewährt wurde. Bei den letzten 150 Fällen handelt es sich um potenzielle Kunden, deren Kreditrisiko die Bank als gering oder hoch einstufen möchte.
- **bankloan_binning.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Informationen zum Finanzstatus und demografischen Hintergrund von 5.000 früheren Kunden enthält.
- **behavior.sav.** In einem klassischen Beispiel (Price als auch Bouffard, 1974) wurden 52 Schüler/Studenten gebeten, die Kombinationen aus 15 Situationen und 15 Verhaltensweisen auf einer 10-Punkte-Skala von 0 = "ausgesprochen angemessen" bis 9 = "ausgesprochen unangemessen" zu bewerten. Die Werte werden über die einzelnen Personen gemittelt und als Unähnlichkeiten verwendet.
- **behavior_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine zweidimensionale Lösung für *behavior.sav*.
- **brakes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik geht, die Scheibenbremsen für Hochleistungsautomobile herstellt. Die Datendatei enthält Messungen des Durchmessers von 16 Scheiben aus 8 Produktionsmaschinen. Der Zieldurchmesser für die Scheiben ist 322 Millimeter.
- **breakfast.sav.** In einer klassischen Studie (Green als auch Rao, 1972) wurden 21 MBA-Studenten der Wharton School mit ihren Lebensgefährten darum gebeten, 15 Frühstücksartikel in der Vorzugsreihenfolge von 1 = "am meisten bevorzugt" bis 15 = "am wenigsten bevorzugt" zu ordnen. Die Bevorzugungen wurden in sechs unterschiedlichen Szenarien erfasst, von "Overall preference" (Allgemein bevorzugt) bis "Snack, with beverage only" (Imbiss, nur mit Getränk).
- **breakfast-overall.sav.** Diese Datei enthält die Daten zu den bevorzugten Frühstücksartikeln, allerdings nur für das erste Szenario, "Overall preference" (Allgemein bevorzugt).
- **broadband_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die die Anzahl der Abonnenten eines Breitband-Service, nach Region geordnet, enthält. Die Datendatei enthält die monatlichen Abonnentenzahlen für 85 Regionen über einen Zeitraum von vier Jahren.
- **broadband_2.sav.** Diese Datendatei stimmt mit *broadband_1.sav* überein, enthält jedoch Daten für weitere drei Monate.

- **car_insurance_claims.sav.** Ein an anderer Stelle (McCullagh als auch Nelder, 1989) vorgestelltes und analysiertes Daten-Set bezieht sich auf Schadensansprüche für Autos. Die durchschnittliche Höhe der Schadensansprüche lässt sich mit Gamma-Verteilung modellieren. Dazu wird eine inverse Verknüpfungsfunktion verwendet, um den Mittelwert der abhängigen Variablen mit einer linearen Kombination aus Alter des Versicherungsnehmers, Fahrzeugtyp und Fahrzeugalter in Bezug zu setzen. Die Anzahl der eingereichten Schadensansprüche kann als Skalierungsgewicht verwendet werden.
- **car_sales.sav.** Diese Datendatei enthält hypothetische Verkaufsschätzer, Listenpreise und physische Spezifikationen für verschiedene Fahrzeugfabrikate und -modelle. Die Listenpreise und physischen Spezifikationen wurden von *edmunds.com* und Hersteller-Websites entnommen.
- **carpet.sav.** In einem beliebten Beispiel möchte (Green als auch Wind, 1973) einen neuen Teppichreiniger vermarkten und dazu den Einfluss von fünf Faktoren auf die Bevorzugung durch den Verbraucher untersuchen: Verpackungsgestaltung, Markenname, Preis, Gütesiegel, *Good Housekeeping* und Geld-zurück-Garantie. Die Verpackungsgestaltung liegt in drei Faktorstufen vor, die sich durch die Position der Auftragebürste unterscheiden. Außerdem gibt es drei Markennamen (*K2R*, *Glory* und *Bissell*), drei Preisstufen sowie je zwei Stufen (Nein oder Ja) für die letzten beiden Faktoren. 10 Kunden stufen 22 Profile ein, die durch diese Faktoren definiert sind. Die Variable *Preference* enthält den Rang der durchschnittlichen Einstufung für die verschiedenen Profile. Ein niedriger Rang bedeutet eine starke Bevorzugung. Diese Variable gibt ein Gesamtmaß der Bevorzugung für die Profile an.
- **carpet_prefs.sav.** Diese Datendatei beruht auf denselben Beispielen, wie für *carpet.sav* beschrieben, enthält jedoch die tatsächlichen Einstufungen durch jeden der 10 Kunden. Die Kunden wurden gebeten, die 22 Produktprofile in der Reihenfolge ihrer Präferenzen einzustufen. Die Variablen *PREF1* bis *PREF22* enthalten die IDs der zugeordneten Profile, wie in *carpet_plan.sav* definiert.
- **catalog.sav.** Diese Datendatei enthält hypothetische monatliche Verkaufszahlen für drei Produkte, die von einem Versandhaus verkauft werden. Daten für fünf mögliche Einflussvariablen wurden ebenfalls aufgenommen.
- **catalog_seasfac.sav.** Diese Datendatei ist mit *catalog.sav* identisch, außer, dass ein Set von saisonalen Faktoren, die mithilfe der Prozedur "Saisonale Zerlegung" berechnet wurden, sowie die zugehörigen Datumsvariablen hinzugefügt wurden.
- **cellular.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Mobiltelefonunternehmens geht, die Kundenabwanderung zu verringern. Scores für die Abwanderungsneigung (von 0 bis 100) werden auf die Kunden angewendet. Kunden mit einem Score von 50 oder höher streben vermutlich einen Anbieterwechsel an.
- **ceramics.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Herstellers geht, der ermitteln möchte, ob ein neue, hochwertige Keramiklegierung eine größere Hitzebeständigkeit aufweist als eine Standardlegierung. Jeder Fall entspricht einem Test einer der Legierungen; die Temperatur, bei der das Keramikwälzlager versagte, wurde erfasst.
- **cereal.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Umfrage geht, bei der 880 Personen nach ihren Frühstückgewohnheiten befragt wurden. Außerdem wurden Alter, Geschlecht, Familienstand und Vorliegen bzw. Nichtvorliegen eines aktiven Lebensstils (auf der Grundlage von mindestens zwei Trainingseinheiten pro Woche) erfasst. Jeder Fall entspricht einem Teilnehmer.

- **clothing_defects.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Bekleidungsfabrik geht. Aus jeder in der Fabrik produzierten Charge entnehmen die Kontrolleure eine Stichprobe an Bekleidungsartikeln und zählen die Anzahl der Bekleidungsartikel die inakzeptabel sind.
- **coffee.sav.** Diese Datendatei enthält Daten zum wahrgenommenen Image von sechs Eiskaffeemarken (Kennedy, Riquier, als auch Sharp, 1996). Bei den 23 Attributen des Eiskaffee-Image sollten die Teilnehmer jeweils alle Marken auswählen, die durch dieses Attribut beschrieben werden. Die sechs Marken werden als "AA", "BB", "CC", "DD", "EE" und "FF" bezeichnet, um Vertraulichkeit zu gewährleisten.
- **contacts.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Kontaktlisten einer Gruppe von Vertretern geht, die Computer an Unternehmen verkaufen. Die einzelnen Kontaktpersonen werden anhand der Abteilung, in der sie in ihrem Unternehmen arbeiten und anhand ihrer Stellung in der Unternehmenshierarchie in Kategorien eingeteilt. Außerdem werden der Betrag des letzten Verkaufs, die Zeit seit dem letzten Verkauf und die Größe des Unternehmens, in dem die Kontaktperson arbeitet, aufgezeichnet.
- **creditpromo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Kaufhauses geht, die Wirksamkeit einer kürzlich durchgeführten Kreditkarten-Werbeaktion einzuschätzen. Dazu wurden 500 Karteninhaber nach dem Zufallsprinzip ausgewählt. Die Hälfte erhielt eine Werbebeilage, die einen reduzierten Zinssatz für Einkäufe in den nächsten drei Monaten ankündigte. Die andere Hälfte erhielt eine Standard-Werbebeilage.
- **customer_dbase.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, das die Informationen in seinem Data Warehouse nutzen möchte, um spezielle Angebote für Kunden zu erstellen, die mit der größten Wahrscheinlichkeit darauf ansprechen. Nach dem Zufallsprinzip wurde eine Untergruppe des Kundenstamms ausgewählt. Diese Gruppe erhielt die speziellen Angebote und die Reaktionen wurden aufgezeichnet.
- **customers_model.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, auf die sich eine Marketingkampagne richtete. Zu diesen Daten gehören demografische Informationen, eine Übersicht über die bisherigen Einkäufe und die Angabe ob die einzelnen Personen auf die Kampagne ansprachen oder nicht. Jeder Fall entspricht einer Einzelperson.
- **customers_new.sav.** Diese Datei enthält hypothetische Daten zu Einzelpersonen, die potenzielle Kandidaten für Marketingkampagnen sind. Zu diesen Daten gehören demografische Informationen und eine Übersicht über die bisherigen Einkäufe für jede Person. Jeder Fall entspricht einer Einzelperson.
- **debate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die gepaarte Antworten auf eine Umfrage unter den Zuhörern einer politischen Debatte enthält (Antworten vor und nach der Debatte). Jeder Fall entspricht einem Befragten.
- **debate_aggregate.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der die Antworten aus *debate.sav* aggregiert wurden. Jeder Fall entspricht einer Kreuzklassifikation der bevorzugten Politiker vor und nach der Debatte.
- **demo.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Kundendatenbank geht, die zum Zwecke der Zusendung monatlicher Angebote erworben wurde. Neben verschiedenen demografischen Informationen ist erfasst, ob der Kunde auf das Angebot geantwortet hat.

- **demo_cs_1.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den ersten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einer anderen Stadt. Außerdem sind IDs für Region, Provinz, Landkreis und Stadt erfasst.
- **demo_cs_2.sav.** Hierbei handelt es sich um eine hypothetische Datendatei für den zweiten Schritt eines Unternehmens, das eine Datenbank mit Umfrageinformationen zusammenstellen möchte. Jeder Fall entspricht einem anderen Stadtteil aus den im ersten Schritt ausgewählten Städten. Außerdem sind IDs für Region, Provinz, Landkreis, Stadt, Stadtteil und Wohneinheit erfasst. Die Informationen zur Stichprobenziehung aus den ersten beiden Stufen des Stichprobenplans sind ebenfalls enthalten.
- **demo_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfrageinformationen enthält die mit einem komplexen Stichprobenplan erfasst wurden. Jeder Fall entspricht einer anderen Wohneinheit. Es sind verschiedene Informationen zum demografischen Hintergrund und zur Stichprobenziehung erfasst.
- **dietstudy.sav.** Diese hypothetische Datendatei enthält die Ergebnisse einer Studie der "Stillman-Diät" (Rickman, Mitchell, Dingman, als auch Dalen, 1974). Jeder Fall entspricht einem Teilnehmer und enthält dessen Gewicht vor und nach der Diät in amerikanischen Pfund sowie mehrere Messungen des Triglyceridspiegels (in mg/100 ml).
- **dischargedata.sav.** Hierbei handelt es sich um eine Datendatei zum Thema *Seasonal Patterns of Winnipeg Hospital Use*, (Menec , Roos, Nowicki, MacWilliam, Finlayson , als auch Black, 1999) (Saisonale Muster der Belegung im Krankenhaus von Winnipeg) vom Manitoba Centre for Health Policy.
- **dvdplayer.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Entwicklung eines neuen DVD-Spielers geht. Mithilfe eines Prototyps hat das Marketing-Team Zielgruppendaten erfasst. Jeder Fall entspricht einem befragten Benutzer und enthält demografische Daten zu dem Benutzer sowie dessen Antworten auf Fragen zum Prototyp.
- **flying.sav.** Diese Datendatei enthält die Flugmeilen zwischen zehn Städten in den USA.
- **german_credit.sav.** Diese Daten sind aus dem Daten-Set "German credit" im Repository of Machine Learning Databases (Blake als auch Merz, 1998) an der Universität von Kalifornien in Irvine entnommen.
- **grocery_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *grocery_coupons.sav*, wobei die wöchentlichen Einkäufe zusammengefasst sind, sodass jeder Fall einem anderen Kunden entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und der verzeichnete ausgegebene Betrag ist nun die Summe der Beträge, die in den vier Wochen der Studie ausgegeben wurden.
- **grocery_coupons.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die Umfragedaten enthält, die von einer Lebensmittelkette erfasst wurden, die sich für die Kaufgewohnheiten ihrer Kunden interessiert. Jeder Kunde wird über vier Wochen beobachtet, und jeder Fall entspricht einer Kundenwoche und enthält Informationen zu den Geschäften, in denen der Kunde einkauft sowie zu anderen Merkmalen, beispielsweise welcher Betrag in der betreffenden Woche für Lebensmittel ausgegeben wurde.
- **guttman.sav.** Bell (Bell, 1961) legte eine Tabelle zur Darstellung möglicher sozialer Gruppen vor. Guttman (Guttman, 1968) verwendete einen Teil dieser Tabelle, bei der fünf Variablen, die Aspekte beschreiben, wie soziale Interaktion, das Gefühl der Gruppenzugehörigkeit, die

physische Nähe der Mitglieder und die Formalität der Beziehung, mit sieben theoretischen sozialen Gruppen gekreuzt wurden: “crowds” (Menschenmassen, beispielsweise die Zuschauer eines Fußballspiels), “audience” (Zuhörerschaften, beispielsweise die Personen im Theater oder bei einer Vorlesung), “public” (Öffentlichkeit, beispielsweise Zeitungsleser oder Fernsehzuschauer), “mobs” (Mobs, wie Menschenmassen, jedoch mit wesentlich stärkerer Interaktion), “primary groups” (Primärgruppen, vertraulich), “secondary groups” (Sekundärgruppen, freiwillig) und “modern community” (die moderne Gesellschaft, ein lockerer Zusammenschluss, der aus einer engen physischen Nähe und dem Bedarf an spezialisierten Dienstleistungen entsteht).

- **healthplans.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Versicherungsgruppe geht, vier verschiedene Pläne zur Gesundheitsvorsorge für Kleinbetriebe zu evaluieren. Zwölf Inhaber von Kleinbetrieben (Arbeitgeber) wurden gebeten, die Pläne danach in eine Rangfolge zu bringen, wie gern sie sie ihren Mitarbeitern anbieten würden. Jeder Fall entspricht einem Arbeitgeber und enthält die Reaktionen auf die einzelnen Pläne.
- **health_funding.sav.** Hierbei handelt es sich um eine hypothetische Datei, die Daten zur Finanzierung des Gesundheitswesens (Betrag pro 100 Personen), Krankheitsraten (Rate pro 10.000 Personen der Bevölkerung) und Besuche bei medizinischen Einrichtungen/Ärzten (Rate pro 10.000 Personen der Bevölkerung) enthält. Jeder Fall entspricht einer anderen Stadt.
- **hivassay.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Bemühungen eines pharmazeutischen Labors, einen Schnelltest zur Erkennung von HIV-Infektionen zu entwickeln. Die Ergebnisse des Tests sind acht kräftiger werdende Rotschattierungen, wobei kräftigeren Schattierungen auf eine höhere Infektionswahrscheinlichkeit hindeuten. Bei 2.000 Blutproben, von denen die Hälfte mit HIV infiziert war, wurde ein Labortest durchgeführt.
- **hourlywagedata.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zum Stundenlohn von Pflegepersonal in Praxen und Krankenhäusern mit unterschiedlich langer Berufserfahrung.
- **insure.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um eine Versicherungsgesellschaft geht, die die Risikofaktoren untersucht, die darauf hinweisen, ob ein Kunde die Leistungen einer mit einer Laufzeit von 10 Jahren abgeschlossenen Lebensversicherung in Anspruch nehmen wird. Jeder Fall in der Datendatei entspricht einem Paar von Verträgen, je einer mit Leistungsforderung und der andere ohne, wobei die beiden Versicherungsnehmer in Alter und Geschlecht übereinstimmen.
- **judges.sav.** Hierbei handelt es sich um eine hypothetische Datendatei mit den Wertungen von ausgebildeten Kampfrichtern (sowie eines Sportliebhabers) zu 300 Kunstturnleistungen. Jede Zeile stellt eine Leistung dar; die Kampfrichter bewerteten jeweils dieselben Leistungen.
- **kinship_dat.sav.** Rosenberg und Kim (Rosenberg als auch Kim, 1975) haben 15 Bezeichnungen für den Verwandtschaftsgrad untersucht (Tante, Bruder, Cousin, Tochter, Vater, Enkelin, Großvater, Großmutter, Enkel, Mutter, Neffe, Nichte, Schwester, Sohn, Onkel). Die beiden Analytiker baten vier Gruppen von College-Studenten (zwei weibliche und zwei männliche Gruppen), diese Bezeichnungen auf der Grundlage der Ähnlichkeiten zu sortieren. Zwei Gruppen (eine weibliche und eine männliche Gruppe) wurden gebeten, die Bezeichnungen zweimal zu sortieren; die zweite Sortierung sollte dabei nach einem anderen Kriterium erfolgen als die erste. So wurden insgesamt sechs “Quellen” erzielt. Jede Quelle entspricht einer Ähnlichkeitsmatrix mit 15×15 Elementen. Die Anzahl der Zellen ist dabei gleich der

Anzahl der Personen in einer Quelle minus der Anzahl der gemeinsamen Platzierungen der Objekte in dieser Quelle.

- **kinship_ini.sav.** Diese Datendatei enthält eine Ausgangskonfiguration für eine dreidimensionale Lösung für *kinship_dat.sav*.
- **kinship_var.sav.** Diese Datendatei enthält die unabhängigen Variablen *gender* (Geschlecht), *gener* (Generation) und *degree* (Verwandtschaftsgrad), die zur Interpretation der Dimensionen einer Lösung für *kinship_dat.sav* verwendet werden können. Insbesondere können sie verwendet werden, um den Lösungsraum auf eine lineare Kombination dieser Variablen zu beschränken.
- **mailresponse.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines Bekleidungsherstellers geht, der ermitteln möchte, ob die Verwendung von Briefsendungen für das Direktmarketing zu schnelleren Antworten führt als Postwurfsendungen. Die Mitarbeiter in der Bestellannahme erfassen, wie vielen Wochen nach der Postsendung die einzelnen Bestellungen aufgegeben wurden.
- **marketvalues.sav.** Diese Datendatei betrifft Hausverkäufe in einem Neubaugebiet in Algonquin, Illinois, in den Jahren 1999–2000. Diese Verkäufe sind in Grundbucheinträgen dokumentiert.
- **mutualfund.sav.** Diese Datendatei betrifft Aktienmarktdaten für verschiedene Technologieaktien, die in im Index S&P 500 verzeichnet sind. Jeder Fall entspricht einem Unternehmen.
- **nhis2000_subset.sav.** Die "National Health Interview Survey (NHIS)" ist eine große, bevölkerungsbezogene Umfrage in unter der US-amerikanischen Zivilbevölkerung. Es werden persönliche Interviews in einer landesweit repräsentativen Stichprobe von Haushalten durchgeführt. Für die Mitglieder jedes Haushalts werden demografische Informationen und Beobachtungen zum Gesundheitsverhalten und Gesundheitsstatus eingeholt. Diese Datendatei enthält eine Teilmenge der Informationen aus der Umfrage des Jahres 2000. National Center for Health Statistics. National Health Interview Survey, 2000. Datendatei und Dokumentation öffentlich zugänglich. *ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/*. Zugriff erfolgte 2003.
- **ozone.sav.** Die Daten enthalten 330 Beobachtungen zu sechs meteorologischen Variablen zur Vorhersage der Ozonkonzentration aus den übrigen Variablen. Bei früheren Untersuchungen (Breiman als auch Friedman, 1985), (Hastie als auch Tibshirani, 1990) fanden Wissenschaftler einige Nichtlinearitäten unter diesen Variablen, die die Standardverfahren bei der Regression behindern.
- **pain_medication.sav.** Diese hypothetische Datendatei enthält die Ergebnisse eines klinischen Tests für ein entzündungshemmendes Medikament zur Schmerzbehandlung bei chronischer Arthritis. Von besonderem Interesse ist die Zeitdauer, bis die Wirkung des Medikaments einsetzt und wie es im Vergleich mit bestehenden Medikamenten abschneidet.
- **patient_los.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen zu Patienten, die wegen des Verdachts auf Herzinfarkt in das Krankenhaus eingeliefert wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.

- **patlos_sample.sav.** Diese hypothetische Datendatei enthält die Behandlungsaufzeichnungen für eine Stichprobe von Patienten, denen während der Behandlung eines Herzinfarkts Thrombolytika verabreicht wurden. Jeder Fall entspricht einem Patienten und enthält diverse Variablen in Bezug auf den Krankenhausaufenthalt.
- **polishing.sav.** Hierbei handelt es sich um die Datendatei "Nambeware Polishing Times" aus der Data and Story Library. Sie bezieht sich auf die Bemühungen eines Herstellers von Metallgeschirr (Nambe Mills, Santa Fe, New Mexico) zur zeitlichen Planung seiner Produktion. Jeder Fall entspricht einem anderen Artikel in der Produktpalette. Für jeden Artikel sind Durchmesser, Polierzeit, Preis und Produkttyp erfasst.
- **poll_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bemühungen geht, die öffentliche Unterstützung für einen Gesetzentwurf zu ermitteln, bevor er im Parlament eingebracht wird. Die Fälle entsprechen registrierten Wählern. Für jeden Fall sind County, Gemeinde und Wohnviertel des Wählers erfasst.
- **poll_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *poll_cs.sav* aufgeführten Wähler. Die Stichprobe wurde gemäß dem in der Plandatei *poll.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Beachten Sie jedoch Folgendes: Da im Stichprobenplan die PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*poll_jointprob.sav*). Die zusätzlichen Variablen zum demografischen Hintergrund der Wähler und ihrer Meinung zum vorgeschlagenen Gesetzentwurf wurden nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.
- **property_assess.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen Bezirk (County) zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien, die im vergangenen Jahr in dem betreffenden County verkauft wurden. Jeder Fall in der Datendatei enthält die Gemeinde, in der sich die Immobilie befindet, den Bewerter, der die Immobilie besichtigt hat, die seit dieser Bewertung verstrichene Zeit, den zu diesem Zeitpunkt ermittelten Wert sowie den Verkaufswert der Immobilie.
- **property_assess_cs.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, in der es um die Bemühungen eines für einen US-Bundesstaat zuständigen Immobilienbewerters geht, trotz eingeschränkter Ressourcen die Einschätzungen des Werts von Immobilien auf dem aktuellsten Stand zu halten. Die Fälle entsprechen den Immobilien in dem betreffenden Bundesstaat. Jeder Fall in der Datendatei enthält das County, die Gemeinde und das Wohnviertel, in dem sich die Immobilie befindet, die seit der letzten Bewertung verstrichene Zeit sowie zu diesem Zeitpunkt ermittelten Wert.
- **property_assess_cs_sample.sav.** Diese hypothetische Datendatei enthält eine Stichprobe der in *property_assess_cs.sav* aufgeführten Immobilien. Die Stichprobe wurde gemäß dem in der Plandatei *property_assess.csplan* angegebenen Stichprobenplan gezogen und in dieser Datendatei sind die Einschlusswahrscheinlichkeiten und Stichprobengewichtungen erfasst. Die zusätzliche Variable *Current value* (Aktueller Wert) wurde nach der Ziehung der Stichprobe erfasst und zur Datendatei hinzugefügt.

- **recidivism.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem frühren Straftäter und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Zeit bis zu seiner zweiten Festnahme, sofern diese innerhalb von zwei Jahren nach der ersten Festnahme erfolgte.
- **recidivism_cs_sample.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Strafverfolgungsbehörde geht, einen Einblick in die Rückfallraten in ihrem Zuständigkeitsbereich zu gewinnen. Jeder Fall entspricht einem früheren Straftäter, der im Juni 2003 erstmals aus der Haft entlassen wurde, und erfasst Daten zu dessen demografischen Hintergrund, einige Details zu seinem ersten Verbrechen sowie die Daten zu seiner zweiten Festnahme, sofern diese bis Ende Juni 2006 erfolgte. Die Straftäter wurden aus per Stichprobenziehung ermittelten Polizeidirektionen ausgewählt (gemäß dem in *recidivism_cs.csplan* angegebenen Stichprobenplan). Da hierbei eine PPS-Methode (PPS: probability proportional to size; Wahrscheinlichkeit proportional zur Größe) verwendet wird, gibt es außerdem eine Datei mit den gemeinsamen Auswahlwahrscheinlichkeiten (*recidivism_cs_jointprob.sav*).
- **salesperformance.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung von zwei neuen Verkaufsschulungen geht. 60 Mitarbeiter, die in drei Gruppen unterteilt sind, erhalten jeweils eine Standardschulung. Zusätzlich erhält Gruppe 2 eine technische Schulung und Gruppe 3 eine Praxisschulung. Die einzelnen Mitarbeiter wurden am Ende der Schulung einem Test unterzogen und die erzielten Punkte wurden erfasst. Jeder Fall in der Datendatei stellt einen Lehrgangsteilnehmer dar und enthält die Gruppe, der der Lehrgangsteilnehmer zugeteilt wurde sowie die von ihm in der Prüfung erreichte Punktzahl.
- **satisf.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Zufriedenheitsumfrage, die von einem Einzelhandelsunternehmen in 4 Filialen durchgeführt wurde. Insgesamt wurden 582 Kunden befragt. Jeder Fall gibt die Antworten eines einzelnen Kunden wieder.
- **screws.sav.** Diese Datendatei enthält Informationen zu den Eigenschaften von Schrauben, Bolzen, Muttern und Reißnägeln (Hartigan, 1975).
- **shampoo_ph.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Qualitätskontrolle in einer Fabrik für Haarpflegeprodukte geht. In regelmäßigen Zeitabständen werden Messwerte von sechs separaten Ausgangschargen erhoben und ihr pH-Wert erfasst. Der Zielbereich ist 4,5–5,5.
- **ships.sav.** Ein an anderer Stelle (McCullagh et al., 1989) vorgestelltes und analysiertes Daten-Set bezieht sich auf die durch Wellen verursachten Schäden an Frachtschiffen. Die Vorfallshäufigkeiten können unter Angabe von Schiffstyp, Konstruktionszeitraum und Betriebszeitraum gemäß einer Poisson-Rate modelliert werden. Das Aggregat der Betriebsmonate für jede Zelle der durch die Kreuzklassifizierung der Faktoren gebildeten Tabelle gibt die Werte für die Risikoanfälligkeit an.
- **site.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Unternehmens geht, neue Standorte für die betriebliche Expansion auszuwählen. Das Unternehmen beauftragte zwei Berater unabhängig voneinander mit der Bewertung der Standorte. Neben einem umfassenden Bericht gaben die Berater auch eine zusammenfassende Wertung für jeden Standort als “good” (gut) “fair” (mittelmäßig) oder “poor” (schlecht) ab.

- **siteratings.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Betatests der neuen Website eines E-Commerce-Unternehmens geht. Jeder Fall entspricht einem Beta-Tester, der die Brauchbarkeit der Website auf einer Skala von 0 bis 20 bewertete.
- **smokers.sav.** Diese Datendatei wurde aus der Umfrage "National Household Survey of Drug Abuse" aus dem Jahr 1998 abstrahiert und stellt eine Wahrscheinlichkeitsstichprobe US-amerikanischer Haushalte dar. Daher sollte der erste Schritt bei der Analyse dieser Datendatei darin bestehen, die Daten entsprechend den Bevölkerungstrends zu gewichten.
- **smoking.sav.** Hierbei handelt es sich um eine von Greenacre (Greenacre , 1984) vorgestellte hypothetische Tabelle. Die relevante Tabelle wird durch eine Kreuztabelle der Rauchgewohnheiten und der Berufskategorie gebildet. Die Variable *Berufsgruppe* enthält die Berufskategorien *Senior Manager*, *Junior Manager*, *Angestellter mit Erfahrung*, *Angestellter ohne Erfahrung* und *Sekretariat* sowie die Kategorie *National Average*, die als Ergänzung der Analyse dienen kann. Die Variable *Rauchen* enthält die Rauchgewohnheiten *Nichtraucher*, *Leicht*, *Mittel* und *Stark* sowie die Kategorien *No Alcohol* und *Alcohol*, die als Ergänzung der Analyse dienen können.
- **storebrand.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen einer Verkaufsleiterin in einem Lebensmittelmarkt geht, die die Verkaufszahlen des Waschmittels der Eigenmarke gegenüber den anderen Marken steigern möchte. Sie erarbeitet eine Werbeaktion im Geschäft und spricht an der Kasse mit Kunden. Jeder Fall entspricht einem Kunden.
- **stores.sav.** Diese Datendatei enthält hypothetische monatliche Marktanteilsdaten für zwei konkurrierende Lebensmittelgeschäfte. Jeder Fall entspricht den Marktanteilsdaten für einen bestimmten Monat.
- **stroke_clean.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozeduren in der Option "Data Preparation" bereinigt wurde.
- **stroke_invalid.sav.** Diese hypothetische Datendatei enthält den ursprünglichen Zustand einer medizinischen Datenbank, der mehrere Dateneingabefehler aufweist.
- **stroke_survival.** In dieser hypothetischen Datendatei geht es um die Überlebenszeiten von Patienten, die nach einem Rehabilitationsprogramm wegen eines ischämischen Schlaganfalls mit einer Reihe von Problemen zu kämpfen haben. Nach dem Schlaganfall werden das Auftreten von Herzinfarkt, ischämischem Schlaganfall und hämorrhagischem Schlaganfall sowie der Zeitpunkt des Ereignisses aufgezeichnet. Die Stichprobe ist auf der linken Seite abgeschnitten, da sie nur Patienten enthält, die bis zum Ende des Rehabilitationprogramms, das nach dem Schlaganfall durchgeführt wurde, überlebten.
- **stroke_valid.sav.** Diese hypothetische Datendatei enthält den Zustand einer medizinischen Datenbank, nachdem diese mithilfe der Prozedur "Daten validieren" überprüft wurde. Sie enthält immer noch potenziell anomale Fälle.
- **tastetest.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Bewertung der Auswirkungen der Mulchfarbe auf den Geschmack von Pflanzenprodukten geht. Der Geschmack von Erdbeeren, die in rotem, blauem und schwarzem Rindenmulch gezogen wurden, wurde von Testpersonen auf einer ordinalen Skala (weit unter bis weit über dem Durchschnitt) bewertet. Jeder Fall entspricht einem Geschmackstester.

- **telco.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Telekommunikationsunternehmens geht, die Kundenabwanderung zu verringern. Jeder Fall entspricht einem Kunden und enthält verschiedene Informationen zum demografischen Hintergrund und zur Servicenutzung.
- **telco_extra.sav.** Diese Datendatei ähnelt der Datei *telco.sav*, allerdings wurden die Variablen "tenure" und die Log-transformierten Variablen zu den Kundenausgaben entfernt und durch standardisierte Log-transformierte Variablen ersetzt.
- **telco_missing.sav.** Diese Datendatei entspricht der Datei *telco_mva_complete.sav*, allerdings wurde ein Teil der Daten durch fehlende Werte ersetzt.
- **telco_mva_complete.sav.** Bei dieser Datendatei handelt es sich um eine Teilmenge der Datendatei *telco.sav*, allerdings mit anderen Variablennamen.
- **testmarket.sav.** Diese hypothetische Datendatei bezieht sich auf die Pläne einer Fast-Food-Kette, einen neuen Artikel in ihr Menü aufzunehmen. Es gibt drei mögliche Kampagnen zur Verkaufsförderung für das neue Produkt. Daher wird der neue Artikel in Filialen in mehreren zufällig ausgewählten Märkten eingeführt. An jedem Standort wird eine andere Form der Verkaufsförderung verwendet und die wöchentlichen Verkaufszahlen für das neue Produkt werden für die ersten vier Wochen aufgezeichnet. Jeder Fall entspricht einer Standort-Woche.
- **testmarket_1month.sav.** Bei dieser hypothetischen Datendatei handelt es sich um die Datendatei *testmarket.sav*, wobei die wöchentlichen Verkaufszahlen zusammengefasst sind, sodass jeder Fall einem Standort entspricht. Dadurch entfallen einige der Variablen, die wöchentlichen Änderungen unterworfen waren, und die verzeichneten Verkaufszahlen sind nun die Summe der Verkaufszahlen während der vier Wochen der Studie.
- **tree_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_credit.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält.
- **tree_missing_data.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zu früheren Bankkrediten enthält und eine große Anzahl fehlender Werte aufweist.
- **tree_score_car.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, die demografische Daten sowie Daten zum Kaufpreis von Fahrzeugen enthält.
- **tree_textdata.sav.** Eine einfache Datendatei mit nur zwei Variablen, die vor allem den Standardzustand von Variablen vor der Zuweisung von Messniveau und Wertelabels zeigen soll.
- **tv-survey.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einer Studie, die von einem Fernsehstudio durchgeführt wurde, das überlegt, ob die Laufzeit eines erfolgreichen Programms verlängert werden soll. 906 Personen wurden gefragt, ob sie das Programm unter verschiedenen Bedingungen ansehen würden. Jede Zeile entspricht einem Befragten; jede Spalte entspricht einer Bedingung.
- **ulcer_recurrence.sav.** Diese Datei enthält Teilinformationen aus einer Studie zum Vergleich der Wirksamkeit zweier Therapien zur Vermeidung des Wiederauftretens von Geschwüren. Es stellt ein gutes Beispiel für intervallzensierte Daten dar und wurde an anderer Stelle (Collett, 2003) vorgestellt und analysiert.

- **ulcer_recurrence_recoded.sav.** In dieser Datei sind die Daten aus *ulcer_recurrence.sav* so umstrukturiert, dass das Modell der Ereigniswahrscheinlichkeit für jedes Intervall der Studie berechnet werden kann und nicht nur die Ereigniswahrscheinlichkeit am Ende der Studie. Sie wurde an anderer Stelle (Collett et al., 2003) vorgestellt und analysiert.
- **verd1985.sav.** Diese Datendatei enthält eine Umfrage (Verdegaal, 1985). Die Antworten von 15 Subjekten auf 8 Variablen wurden aufgezeichnet. Die relevanten Variablen sind in drei Sets unterteilt. Set 1 umfasst *alter* und *heirat*, Set 2 besteht aus *pet* und *news* und in Set 3 finden sich *music* und *live*. Die Variable *pet* wird mehrfach nominal skaliert und die Variable *Alter* ordinal. Alle anderen Variablen werden einzeln nominal skaliert.
- **virus.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um die Bemühungen eines Internet-Dienstanbieters geht, der die Auswirkungen eines Virus auf seine Netzwerke ermitteln möchte. Dabei wurde vom Moment der Virusentdeckung bis zu dem Zeitpunkt, zu dem die Virusinfektion unter Kontrolle war, der (ungefähre) prozentuale Anteil infizierter E-Mail in den Netzwerken erfasst.
- **waittimes.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu den Wartezeiten für Kunden bei drei verschiedenen Filialen einer Bank. Jeder Fall entspricht einem Kunden und zeichnet die Wartezeit und die Filiale.
- **webusability.sav.** Hierbei handelt es sich um eine hypothetische Datendatei, bei der es um Tests zur Benutzerfreundlichkeit eines neuen Internetgeschäfts geht. Jeder Fall entspricht einer von fünf Testpersonen, die die Benutzerfreundlichkeit bewerten und gibt für sechs separate Aufgaben an, ob die Testperson sie erfolgreich ausführen könnte.
- **wheeze_steubenville.sav.** Hierbei handelt es sich um eine Teilmenge der Daten aus einer Langzeitstudie zu den gesundheitlichen Auswirkungen der Luftverschmutzung auf Kinder (Ware, Dockery, Spiro III, Speizer, als auch Ferris Jr., 1984). Die Daten enthalten wiederholte binäre Messungen des Keuchens von Kindern aus Steubenville, Ohio, im Alter von 7, 8, 9 und 10 Jahren sowie eine unveränderlichen Angabe, ob die Mutter im ersten Jahr der Studie rauchte oder nicht.
- **workprog.sav.** Hierbei handelt es sich um eine hypothetische Datendatei zu einem Arbeitsprogramm der Regierung, das versucht, benachteiligten Personen bessere Arbeitsplätze zu verschaffen. Eine Stichprobe potenzieller Programmteilnehmer wurde beobachtet. Von diesen Personen wurden nach dem Zufallsprinzip einige für die Teilnahme an dem Programm ausgewählt. Jeder Fall entspricht einem Programmteilnehmer.

Bibliografie

Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

Blake, C. L., als auch C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at http://www.ics.uci.edu/~mlearn/MLRepository.html.

Box, G. E. P., G. M. Jenkins, als auch G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*, 3rd (Hg.). Englewood Cliffs, N.J.: Prentice Hall.

Breiman, L., als auch J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580–598.

Collett, D. 2003. *Modelling survival data in medical research*, 2 (Hg.). Boca Raton: Chapman & Hall/CRC.

Gardner, E. S. 1985. Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1–28.

Green, P. E., als auch V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

Green, P. E., als auch Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

Greenacre , M. J. 1984. *Theory and applications of correspondence analysis*. London: Academic Press.

Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469–506.

Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

Hastie, T., als auch R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

Kennedy, R., C. Riquier, als auch B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56–70.

McCullagh, P., als auch J. A. Nelder. 1989. *Generalized Linear Models*, 2nd (Hg.). London: Chapman & Hall.

Menec , V., N. Roos, D. Nowicki, L. MacWilliam, G. Finlayson , als auch C. Black. 1999. *Seasonal Patterns of Winnipeg Hospital Use*. : Manitoba Centre for Health Policy.

Pena, D., G. C. Tiao, als auch R. S. Tsay (Hgg.). 2001. *A course in time series analysis*. New York: John Wiley and Sons.

Price, R. H., als auch D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579–586.

Rickman, R., N. Mitchell, J. Dingman, als auch J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, 54–58.

Rosenberg, S., als auch M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489–502.

Van der Ham, T., J. J. Meulman, D. C. Van Strien, als auch H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363–368.

Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in niederländischer Sprache)*. Leiden: Department of Data Theory, Universität Leiden.

Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, als auch B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366–374.

Index

ACF
  Diagramme für reine ARIMA-Prozesse, 104
  in Zeitreihenmodell zuweisen, 31, 33
  in Zeitreihenmodellierung, 17, 19
Additiver Ausreißer, 103
  in Zeitreihenmodellierung, 9, 16
Additiver Bereich (Ausreißer), 103
  in Zeitreihenmodellierung, 9, 16
Anpassungsgüte des Modells (Tabelle)
  in Zeitreihenmodell zuweisen, 63
Anpassungswerte
  in Zeitreihenmodell zuweisen, 33
  in Zeitreihenmodellierung, 19, 73
ARIMA-Modelle, 7
  Ausreißer, 16
  autoregressive Ordnungen, 13
  Differenzierungsordnungen, 13
  Konstante, 13
  Ordnungen des gleitenden Durchschnitts, 13
  saisonale Ordnungen, 13
  Übertragungsfunktionen, 14
ARIMA-Modellparameter (Tabelle)
  in Zeitreihenmodellierung, 76
Ausreißer
  ARIMA-Modelle, 16
  Definitionen, 103
  Expert Modeler, 9, 70
Autokorrelationsfunktion
  Diagramme für reine ARIMA-Prozesse, 104
  in Zeitreihenmodell zuweisen, 31, 33
  in Zeitreihenmodellierung, 17, 19
Autoregression
  ARIMA-Modelle, 13

Beispieldateien
  Speicherort, 108
Box-Ljung-Statistik
  in Zeitreihenmodell zuweisen, 31
  in Zeitreihenmodellierung, 17, 75

Differenztransformation
  ARIMA-Modelle, 13

Einfaches Modell mit exponentiellem Glätten, 10
Einfaches saisonales Modell mit exponentiellem Glätten, 10
Ereignisse, 9
  in Zeitreihenmodellierung, 8
Expert Modeler, 7, 49
  Ausreißer, 9, 70
  Modellraum einschränken, 8, 52

Fehlende Werte
  in Zeitreihenmodell zuweisen, 38
  in Zeitreihenmodellierung, 24

Gedämpftes Modell mit exponentiellem Glätten, 10
Gleitender Durchschnitt
  ARIMA-Modelle, 13
Güte der Anpassung
  Definitionen, 101
  in Zeitreihenmodell zuweisen, 31, 33, 63
  in Zeitreihenmodellierung, 17, 19, 55

Harmonische Analyse, 44
Historische Daten
  in Zeitreihenmodell zuweisen, 33
  in Zeitreihenmodellierung, 19
Historische Periode, 2

Im Niveau verschobener Ausreißer, 103
  in Zeitreihenmodellierung, 9, 16
Innovativer Ausreißer, 103
  in Zeitreihenmodellierung, 9, 16
Integration
  ARIMA-Modelle, 13

Konfidenzintervalle
  in Zeitreihenmodell zuweisen, 33, 38
  in Zeitreihenmodellierung, 19, 24

Log-Transformation
  in Zeitreihenmodellierung, 10, 12, 14
Lokaler Trend (Ausreißer), 103
  in Zeitreihenmodellierung, 9, 16

MAE, 101
  in Zeitreihenmodell zuweisen, 31, 33
  in Zeitreihenmodellierung, 17, 19
MAPE, 101
  in Zeitreihenmodell zuweisen, 31, 33, 64
  in Zeitreihenmodellierung, 17, 19, 56
MaxAE, 101
  in Zeitreihenmodell zuweisen, 31, 33
  in Zeitreihenmodellierung, 17, 19
MaxAPE, 101
  in Zeitreihenmodell zuweisen, 31, 33, 64
  in Zeitreihenmodellierung, 17, 19, 56
Maximaler absoluter Fehler, 101
  in Zeitreihenmodell zuweisen, 31, 33
  in Zeitreihenmodellierung, 17, 19

Maximaler absoluter Fehler in Prozent, 101
in Zeitreihenmodell zuweisen, 31, 33, 64
in Zeitreihenmodellierung, 17, 19, 56
Mittlerer absoluter Fehler, 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19
Mittlerer absoluter Fehler in Prozent, 101
in Zeitreihenmodell zuweisen, 31, 33, 64
in Zeitreihenmodellierung, 17, 19, 56
Modell mit exponentiellem Glätten nach Brown, 10
Modell mit exponentiellem Glätten nach Holt, 10
Modell mit exponentiellem Glätten nach Winters
additiv, 10
multiplikativ, 10
Modellbeschreibung (Tabelle)
in Zeitreihenmodellierung, 74
Modelle
ARIMA, 7, 13
Expert Modeler, 7
exponentielles Glätten, 7, 10
Modelle mit exponentiellem Glätten, 7, 10
Modellnamen
in Zeitreihenmodellierung, 24
Modellparameter
in Zeitreihenmodell zuweisen, 31
in Zeitreihenmodellierung, 17, 72
Modellparameter erneut schätzen
in Zeitreihenmodell zuweisen, 29, 61
Modellstatistik (Tabelle)
in Zeitreihenmodellierung, 75

Normalisiertes BIC (Bayes-Informationskriterium), 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19

PACF
Diagramme für reine ARIMA-Prozesse, 104
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19
Partielle Autokorrelationsfunktion
Diagramme für reine ARIMA-Prozesse, 104
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19
Periodizität
in Zeitreihenmodellierung, 8, 10, 12, 14
Prüffälle, 2

Quadratwurzeltransformation
in Zeitreihenmodellierung, 10, 12, 14

R^2, 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19
R^2 für stationären Teil, 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19, 75
Residuen
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19
RMSE, 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19

Saisonal additiver Ausreißer, 103
in Zeitreihenmodellierung, 9, 16
Saisonale Differenztransformation
ARIMA-Modelle, 13
saisonale Ordnungen
ARIMA-Modelle, 13
Saisonale Zerlegung, 40, 42–43
Annahmen, 40
Berechnen der gleitenden Durchschnitte, 40
Erstellen von Variablen, 42
Modelle, 40
neue Variablen, 93
periodische Datumskomponente, 88
Speichern von neuen Variablen, 42
verwandte Prozeduren, 95
Schätzperiode, 2
in Zeitreihenmodell zuweisen, 29
in Zeitreihenmodellierung, 7, 52
speichern
erneut geschätzte Modelle als XML, 37
Modellspezifikationen als XML, 22
Modellvorhersagen, 22, 37
neue Variablennamen, 22, 37
Spektraldiagramme, 44, 47
Annahmen, 44
bivariate Spektralanalyse, 46
Periodogramm, 98
Spektraldichte, 98
Spektralfenster, 44
verwandte Prozeduren, 100
Zentriertransformation, 46

Transformation mit natürlichem Logarithmus
in Zeitreihenmodellierung, 10, 12, 14
Transienter Ausreißer, 103
in Zeitreihenmodellierung, 9, 16

Übertragungsfunktionen, 14
Nenner-Terme, 14
Ordnung der Differenzen, 14
saisonale Ordnungen, 14
Verzögerung, 14
Zähler-Terme, 14

Validierungsperiode, 2
Variablennamen
in Zeitreihenmodell zuweisen, 37

in Zeitreihenmodellierung, 22
Vorhersagen
in Zeitreihenmodell zuweisen, 31, 33, 85
in Zeitreihenmodellierung, 17, 19, 55
Vorhersageperiode
in Zeitreihenmodell zuweisen, 29, 61, 84
in Zeitreihenmodellierung, 7, 24, 52–53
Vorhersagetabelle
in Zeitreihenmodell zuweisen, 85
in Zeitreihenmodellierung, 59

Wurzel der mittleren Fehler-Quadratsumme, 101
in Zeitreihenmodell zuweisen, 31, 33
in Zeitreihenmodellierung, 17, 19

XML
erneut geschätzte Modelle als XML speichern, 37
Zeitreihenmodelle als XML speichern, 22, 54, 71

Zeitreihenmodell zuweisen, 27, 60, 77
am besten/schlechtesten angepasste Modelle, 35
Anpassungsgüte des Modells (Tabelle), 63
Anpassungswerte, 33
Box-Ljung-Statistik, 31
erneut geschätzte Modelle als XML speichern, 37
Fehlende Werte, 38
Konfidenzintervalle, 33, 38
Modellparameter, 31
Modellparameter erneut schätzen, 29, 61
neue Variablennamen, 37, 64
Partielle Residuen-Autokorrelationsfunktion, 31, 33
Residuen-Autokorrelationsfunktion, 31, 33
Schätzperiode, 29
Statistik für alle Modelle, 31, 33, 63
Statistik für Anpassungsgüte, 31, 33, 63
Vorhersagen, 31, 33, 85
Vorhersagen speichern, 37, 62
Vorhersageperiode, 29, 61, 84
Vorhersagetabelle, 85
Zeitreihenmodellierung, 4
am besten/schlechtesten angepasste Modelle, 21
Anpassungswerte, 19, 73
ARIMA, 7, 12
ARIMA-Modellparameter (Tabelle), 76
Ausreißer, 9, 16, 70
Box-Ljung-Statistik, 17
Ereignisse, 8
Expert Modeler, 7, 49, 66
exponentielles Glätten, 7, 10
Fehlende Werte, 24
Konfidenzintervalle, 19, 24
Modellbeschreibung (Tabelle), 74
Modellnamen, 24
Modellparameter, 17, 72
Modellspezifikationen als XML speichern, 22, 54, 71
Modellstatistik (Tabelle), 75
neue Variablennamen, 22, 58
Partielle Residuen-Autokorrelationsfunktion, 17, 19
Periodizität, 8, 10, 12, 14
Residuen-Autokorrelationsfunktion, 17, 19
Schätzperiode, 7, 52
Statistik für alle Modelle, 17, 19, 55, 57
Statistik für Anpassungsgüte, 17, 19, 55, 75
Übertragungsfunktionen, 14
Vorhersagen, 17, 19, 55
Vorhersagen speichern, 22, 54
Vorhersageperiode, 7, 24, 52–53
Vorhersagetabelle, 59
Zeitreihentransformation, 10, 12, 14